VOL. LIV. NO. 214.] [April, 1945.

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—STUDIES IN THE LOGIC OF CONFIRMATION (II.).

BY CARL G. HEMPEL.

7. *The Prediction-criterion of Confirmation and its Shortcomings.*—We are now in a position to analyze a second conception of confirmation which is reflected in many methodological discussions and which can claim a great deal of plausibility. Its basic idea is very simple: General hypotheses in science as well as in everyday usage are intended to enable us to anticipate future events; hence, it seems reasonable to count any prediction which is borne out by subsequent observation as confirming evidence for the hypothesis on which it is based, and any prediction that fails as disconfirming evidence. To illustrate: Let H_1 be the hypothesis that all metals, when heated, expand; symbolically: '(x) ((Metal (x) . Heated (x)) $\supset$ Exp(x))'. If we are given an observation report to the effect that a certain object a is a metal and is heated, then by means of H_1 we can derive the prediction that a expands. Suppose that this is borne out by observation and described in an additional observation statement. We should then have the total observation report. {Metal(a), Heated(a), Exp.(a)}.[1] This report would be qualified as confirming evidence for H_1 because its last sentence bears out what could be predicted, or derived, from the first two by means of

[1] An (observation) report, it will be recalled, may be represented by a conjunction or by a class of observation sentences; in the latter case, we characterize it by writing the sentences between braces; the quotation marks which normally would be used are, for convenience, assumed to be absorbed by the braces.

H_1; more explicitly: because the last sentence can be derived from the first two in conjunction with H_1.—Now let H_2 be the hypothesis that all swans are white; symbolically: '(x) (Swan $(x) \supset$ White(x))'; and consider the observation report {Swan(a), $\sim$ White(a)}. This report would constitute disconfirming evidence for H_2 because the second of its sentences contradicts (and thus fails to bear out) the prediction 'White(a)' which can be deduced from the first sentence in conjunction with H_2; or, symmetrically, because the first sentence contradicts the consequence '$\sim$ Swan(a)' which can be derived from the second in conjunction with H_2. Obviously, either of these formulations implies that H_2 is incompatible with the given observation report.

These illustrations suggest the following general definition of confirmation as successful prediction:

Prediction-criterion of Confirmation: Let H be a hypothesis, B an observation report, *i.e.* a class of observation sentences. Then

(a) B is said to confirm H if B can be divided into two mutually exclusive subclasses B_1 and B_2 such that B_2 is not empty, and every sentence of B_2 can be logically deduced from B_1 in conjunction with H, but not from B_1 alone.

(b) B is said to disconfirm H if H logically contradicts B.[1]

(c) B is said to be neutral with respect to H if it neither confirms nor disconfirms H.[2]

But while this criterion is quite sound as a statement of sufficient conditions of confirmation for hypotheses of the type illustrated above, it is considerably too narrow to serve as a general definition of confirmation. Generally speaking, this criterion would serve its purpose if all scientific hypotheses could be construed as asserting regular connections of observable features in the subject-matter under investigation; *i.e.* if they all were of

[1] It might seem more natural to stipulate that B disconfirms H if it can be divided into two mutually exclusive classes B_1 and B_2 such that the denial of at least one sentence in B_2 can be deduced from B_1 in conjunction with H; but this condition can be shown to be equivalent to (b) above.

[2] The following quotations from A. J. Ayer's book *Language, Truth and Logic* (London, 1936) formulate in a particularly clear fashion the conception of confirmation as successful prediction (although the two are not explicitly identified by definition): "... the function of an empirical hypothesis is to enable us to anticipate experience. Accordingly, if an observation to which a given proposition is relevant conforms to our expectations, ... that proposition is confirmed" (*loc. cit.* pp. 142-143). "... it is the mark of a genuine factual proposition ... that some experiential propositions can be deduced from it in conjunction with certain premises without being deducible from those other premises alone". (*loc. cit.* p. 26).

the form "Whenever the observable characteristic P is present in an object or a situation, then the observable characteristic Q will also be present." But actually, most scientific hypotheses and laws are not of this simple type; as a rule, they express regular connections of characteristics which are not observable in the sense of direct observability, nor even in a much more liberal sense. Consider, for example, the following hypothesis: "Whenever plane-polarized light of wave length λ traverses a layer of quartz of thickness d, then its plane of polarization is rotated through an angle α which is proportional to $\frac{d}{\lambda}$."—Let us assume that the observational vocabulary, by means of which our observation reports have to be formulated, contains exclusively terms referring to directly observable attributes. Then, since the question of whether a given ray of light is plane-polarized and has the wave length λ cannot be decided by means of direct observation, no observation report of the kind here admitted could include information of this type. This in itself would not be crucial if at least we could assume that the fact that a given ray of light is plane-polarized, etc., could be logically inferred from some possible observation report; for then, from a suitable report of this kind, in conjunction with the given hypothesis, one would be able to predict a rotation of the plane of polarization; and from this prediction, which itself is not yet expressed in exclusively observational terms, one might expect to derive further predictions in the form of genuine observation sentences. But actually, a hypothesis to the effect that a given ray of light is plane-polarized has to be considered as a general hypothesis which entails an unlimited number of observation sentences; thus it cannot be logically inferred from, but at best be confirmed by, a suitable set of observational findings. The logically essential point can best be exhibited by reference to a very simple abstract case: Let us assume that R_1 and R_2 are two relations of a kind accessible to direct observation, and that the field of scientific investigation contains infinitely many objects. Consider now the hypothesis

$$(H) \qquad (x)((y)R_1(x, y) \supset (\mathbf{E}z)R_2(x, z)),$$

i.e.: Whenever an object x stands in R_1 to every object y, then it stands in R_2 to at least one object z.—This simple hypothesis has the following property: However many observation sentences may be given, H does not enable us to derive any new observation sentences from them. Indeed—to state the reason in suggestive though not formally rigorous terms—in order to

make a prediction concerning some specific object a, we should first have to know that a stands in R_1 to every object; and this necessary information clearly cannot be contained in any finite number, however large, of observation sentences, because a finite set of observation sentences can tell us at best for a finite number of objects that a stands in R_1 to them. Thus an observation report, which always involves only a finite number of observation sentences, can never provide a sufficiently broad basis for a prediction by means of H.[1]—Besides, even if we did know that a stood in R_1 to every object, the prediction derivable by means of H would not be an observation sentence; it would assert that a stands in R_2 to *some* object, without specifying which, and where to find it. Thus, H would be an empirical hypothesis, containing, besides purely logical terms, only expressions belonging to the observational vocabulary, and yet the predictions which it renders possible neither start from nor lead to observation reports.

It is, therefore, a considerable over-simplification to say that scientific hypotheses and theories enable us to derive predictions of future experiences from descriptions of past ones. Unquestionably, scientific hypotheses do have a predictive function; but the way in which they perform this function, the manner in which they establish logical connections between observation reports, is logically more complex than a deductive inference. Thus, in the last illustration, the predictive use of H may assume the following form: On the basis of a number of individual tests, which show that a does stand in R_1 to three objects b, c, and d, we may accept the hypothesis that a stands in R_1 to all objects; or, in terms of our formal mode of speech: In view of the observation report $\{R_1(a, b), R_1(a, c), R_1(a, d)\}$, the hypothesis that $(y)R_1(a, y)$ is accepted as confirmed by, though not logically inferable from, that report.[2] This process might be referred to as quasi-induction.[3] From the hypothesis thus established we

[1] To illustrate: a might be an iron object which possibly is a magnet; R_1 might be the relation of attracting; the objects under investigation might be iron objects. Then a finite number of observation reports to the effect that a did attract a particular piece of iron is insufficient to *infer* that a will attract every piece of iron.

[2] Thus, in the illustration given in the preceding footnote, the hypothesis that the object a will attract every piece of iron might be accepted as sufficiently well substantiated by, though by no means derivable from, an observation report to the effect that in tests a did attract the iron objects b, c, and d.

[3] The prefix "quasi" is to contradistinguish the procedure in question from so-called induction, which is usually supposed to be a method of discovering, or arriving at, general regularities on the basis of a finite

can then proceed to derive, by means of H, the prediction that a stands in R_2 to at least one object. This again, as was pointed out above, is not an observation sentence; and indeed no observation sentence can be derived from it; but it can, in turn, be confirmed by a suitable observation sentence, such as '$R_2(a, b)$'. —In other cases, the prediction of actual observation sentences may be possible; thus if the given hypothesis asserts that $(x)((y)R_1(x, y) \supset (z)R_2(x, z))$, then after quasi-inductively accepting, as above, that $(y)R_1(a, y)$, we can derive, by means of the given hypothesis, the sentence that a stands in R_2 to every object, and thence, we can deduce special predictions such as '$R_2(a, b)$', etc., which do have the form of observation sentences.

Thus, the chain of reasoning which leads from given observational findings to the "prediction" of new ones actually involves, besides deductive inferences, certain quasi-inductive steps each of which consists in the acceptance of an intermediate statement on the basis of confirming, but usually not logically conclusive, evidence. In most scientific predictions, this general pattern occurs in multiple re-iteration; an analysis of the predictive use of the hypothesis mentioned above, concerning plane-polarized light, could serve as an illustration. In the present context, however, this general account of the structure of scientific prediction is sufficient: it shows that a general definition of confirmation by reference to successful prediction becomes circular; indeed, in order to make the original formulation of the prediction-criterion of confirmation sufficiently comprehensive, we should have to replace the phrase "can be logically deduced" by "can be obtained by a series of steps of deduction and quasi-induction"; and the definition of "quasi-induction" in the above sense presupposes the concept of confirmation.

Let us note, as a by-product of the preceding consideration, the fact that an adequate analysis of scientific prediction (and analogously, of scientific explanation, and of the testing of empirical hypotheses) requires an analysis of the concept of confirmation. The reason for this fact may be restated in general terms as follows: Scientific laws and theories, as a rule, connect terms which lie on the level of abstract theoretical constructs rather than on that of direct observation; and from observation sentences, no merely deductive logical inference leads

number of instances. In quasi-induction, the hypothesis is not "discovered" but has to be *given* in addition to the observation report; the process consists in the acceptance of the hypothesis if it is deemed sufficiently confirmed by the observation report. *Cf.* also the discussion in section 1c, above.

to statements about those theoretical constructs which are the starting point for scientific predictions; statements about logical constructs, such as "This piece of iron is magnetic" or "Here, a plane-polarized ray of light traverses a quartz crystal" can be confirmed, but not entailed, by observation reports, and thus, even though based on general scientific laws, the "prediction" of new observational findings on the basis of given ones is a process involving confirmation in addition to logical deduction.[1]

8. *Conditions of Adequacy for any Definition of Confirmation.*— The two most customary conceptions of confirmation, which were rendered explicit in Nicod's criterion and in the prediction criterion, have thus been found unsuitable for a general definition of confirmation. Besides this negative result, the preceding analysis has also exhibited certain logical characteristics of scientific prediction, explanation, and testing, and it has led to the establishment of certain standards which an adequate definition of confirmation has to satisfy. These standards include the equivalence condition and the requirement that the definition of confirmation be applicable to hypotheses of any degree of logical complexity, rather than to the simplest type of universal conditional only. An adequate definition of confirmation, however, has to satisfy several further logical requirements, to which we now turn.

First of all, it will be agreed that any sentence which is entailed by—*i.e.* a logical consequence of—a given observation report has to be considered as confirmed by that report: Entailment is a special case of confirmation. Thus, *e.g.*, we want to say that the observation report "*a* is black" confirms the sentence (hypothesis) "*a* is black or grey"; and—to refer to one of the illustrations given in the preceding section—the observation sentence '$R_2(a, b)$' should certainly be confirming evidence for the sentence '$(Ez)R_2(a, z)$'. We are therefore led to the stipulation that any adequate definition of confirmation must insure the fulfilment of the

[1] In the above sketch of the structure of scientific prediction, we have disregarded the fact that in practically every case where a prediction is said to be obtained by means of a certain hypothesis or theory, a considerable mass of auxiliary theories is used in addition; thus, *e.g.* the prediction of observable effects of the deflection of light in the gravitational field of the sun on the basis of the general theory of relativity, requires such auxiliary theories as mechanics and optics. But an explicit consideration of this fact would not affect our result that scientific predictions, even when based on hypotheses or theories of universal form, still are not purely deductive in character, but involve quasi-inductive steps as well.

(8.1) *Entailment condition:* Any sentence which is entailed by an observation report is confirmed by it.[1]

This condition is suggested by the preceding consideration, but of course not proved by it. To make it a standard of adequacy for the definition of confirmation means to lay down the stipulation that a proposed definition of confirmation will be rejected as logically inadequate if it is not constructed in such a way that (8.1) is unconditionally satisfied. An analogous remark applies to the subsequently proposed further standards of adequacy.—

Second, an observation report which confirms certain hypotheses would invariably be qualified as confirming any consequence of those hypotheses. Indeed: any such consequence is but an assertion of all or part of the combined content of the original hypotheses and has therefore to be regarded as confirmed by any evidence which confirms the original hypotheses. This suggests the following condition of adequacy:

(8.2) *Consequence Condition:* If an observation report confirms every one of a class K of sentences, then it also confirms any sentence which is a logical consequence of K.

If (8.2) is satisfied, then the same is true of the following two more special conditions:

(8.21) *Special Consequence Condition:* If an observation report confirms a hypothesis H, then it also confirms every consequence of H.

(8.22) *Equivalence Condition:* If an observation report confirms a hypothesis H, then it also confirms every hypothesis which is logically equivalent with H.

(This follows from (8.21) in view of the fact that equivalent hypotheses are mutual consequences of each other.) Thus, the satisfaction of the consequence condition entails that of our earlier equivalence condition, and the latter loses its status of an independent requirement.

In view of the apparent obviousness of these conditions, it is interesting to note that the definition of confirmation in terms of successful prediction, while satisfying the equivalence condition, would violate the consequence condition. Consider, for example, the formulation of the prediction-criterion given in the earlier

[1] As a consequence of this stipulation, a contradictory observation report, such as {Black(a), $\sim$ Black(a)} confirms every sentence, because it has every sentence as a consequence. Of course, it is possible to exclude the possibility of contradictory observation reports altogether by a slight restriction of the definition of "observation report". There is, however, no important reason to do so.

part of the preceding section. Clearly, if the observational findings B_2 can be predicted on the basis of the findings B_1 by means of the hypothesis H, the same prediction is obtainable by means of any equivalent hypothesis, but not generally by means of a weaker one.

On the other hand, any prediction obtainable by means of H can obviously also be established by means of any hypothesis which is stronger than H, *i.e.* which logically entails H. Thus, while the consequence condition stipulates in effect that whatever confirms a given hypothesis also confirms any weaker hypothesis, the relation of confirmation defined in terms of successful prediction would satisfy the condition that whatever confirms a given hypothesis, also confirms every stronger one.

But is this " converse consequence condition ", as it might be called, not reasonable enough, and should it not even be included among our standards of adequacy for the definition of confirmation ? The second of these two suggestions can be readily disposed of : The adoption of the new condition, in addition to (8.1) and (8.2), would have the consequence that any observation report B would confirm any hypothesis H whatsoever. Thus, *e.g.*, if B is the report " a is a raven " and H is Hooke's law, then, according to (8.1), B confirms the sentence " a is a raven ", hence B would, according to the converse consequence condition, confirm the stronger sentence " a is a raven, and Hooke's law holds " ; and finally, by virtue of (8.2), B would confirm H, which is a consequence of the last sentence. Obviously, the same type of argument can be applied in all other cases.

But is it not true, after all, that very often observational data which confirm a hypothesis H are considered also as confirming a stronger hypothesis ? Is it not true, for example, that those experimental findings which confirm Galileo's law, or Kepler's laws, are considered also as confirming Newton's law of gravitation ? [1] This is indeed the case, but this does not justify the acceptance of the converse entailment condition as a general rule of the logic of confirmation ; for in the cases just mentioned, the weaker hypothesis is connected with the stronger one by a logical bond of a particular kind : it is essentially a substitution instance of the stronger one ; thus, *e.g.*, while the law of gravitation refers to the force obtaining between any two bodies, Galileo's law is a specialization referring to the case where one of

[1] Strictly speaking, Galileo's law and Kepler's laws can be deduced from the law of gravitation only if certain additional hypotheses—including the laws of motion—are presupposed ; but this does not affect the point under discussion.

the bodies is the earth, the other an object near its surface. In the preceding case, however, where Hooke's law was shown to be confirmed by the observation report that *a* is a raven, this situation does not prevail; and here, the rule that whatever confirms a given hypothesis also confirms any stronger one becomes an entirely absurd principle. Thus, the converse consequence condition does not provide a sound general condition of adequacy.[1]

A third condition remains to be stated:[2]

(8.3) *Consistency Condition:* Every logically consistent observation report is logically compatible with the class of all the hypotheses which it confirms.

The two most important implications of this requirement are the following:

(8.31) Unless an observation report is self-contradictory,[3] it does not confirm any hypothesis with which it is not logically compatible.

(8.32) Unless an observation report is self-contradictory, it does not confirm any hypotheses which contradict each other.

The first of these corollaries will readily be accepted; the second, however,—and consequently (8.3) itself—will perhaps be

[1] William Barrett, in a paper entitled "Discussion on Dewey's Logic" (*The Philosophical Review*, vol. 50, 1941, pp. 305 ff., esp. p. 312) raises some questions closely related to what we have called above the consequence condition and the converse consequence condition. In fact, he invokes the latter (without stating it explicitly) in an argument which is designed to show that "not every observation which confirms a sentence need also confirm all its consequences", in other words, that the special consequence condition (8.21) need not always be satisfied. He supports his point by reference to "the simplest case: the sentence 'C' is an abbreviation of 'A.B', and the observation O confirms 'A', *and so* 'C', but is irrelevant to 'B', which is a consequence of 'C'." (Italics mine.)

For reasons contained in the above discussion of the consequence condition and the converse consequence condition, the application of the latter in the case under consideration seems to us unjustifiable, so that the illustration does not prove the author's point; and indeed, there seems to be every reason to preserve the unrestricted validity of the consequence condition. As a matter of fact, Mr. Barrett himself argues that "the degree of confirmation for the consequence of a sentence cannot be less than that of the sentence itself"; this is indeed quite sound; but it is hard to see how the recognition of this principle can be reconciled with a renunciation of the special consequence condition, since the latter may be considered simply as the correlate, for the non-gradated relation of confirmation, of the former principle which is adapted to the concept of degree of confirmation.

[2] For a fourth condition, see n. 1, p. 110.

[3] A contradictory observation report confirms every hypothesis (*cf.* n. 1, p. 103) and is, of course, incompatible with every one of the hypotheses it confirms.

felt to embody a too severe restriction. It might be pointed out, for example, that a finite set of measurements concerning the variation of one physical magnitude, x, with another, y, may conform to, and thus be said to confirm, several different hypotheses as to the particular mathematical function in terms of which the relationship of x and y can be expressed; but such hypotheses are incompatible because to at least one value of x, they will assign different values of y.

No doubt it is possible to liberalize the formal standards of adequacy in line with these considerations. This would amount to dropping (8.3) and (8.32) and retaining only (8.31). One of the effects of this measure would be that when a logically consistent observation report B confirms each of two hypotheses, it does not necessarily confirm their conjunction; for the hypotheses might be mutually incompatible, hence their conjunction self-contradictory; consequently, by (8.31), B could not confirm it.—This consequence is intuitively rather awkward, and one might therefore feel inclined to suggest that while (8.3) should be dropped and (8.31) retained, (8.32) should be replaced by the requirement (8.33): If an observation sentence confirms each of two hypotheses, then it also confirms their conjunction. But it can readily be shown that by virtue of (8.2) this set of conditions entails the fulfilment of (8.32).

If, therefore, the condition (8.3) appears to be too rigorous, the most obvious alternative would seem to lie in replacing (8.3) and its corollaries by the much weaker condition (8.31) alone; and it is an important problem whether an intuitively adequate definition of confirmation can be constructed which satisfies (8.1), (8.2) and (8.31), but not (8.3).—One of the great advantages of a definition which satisfies (8.3) is that it sets a limit, so to speak, to the strength of the hypotheses which can be confirmed by given evidence.[1]

The remainder of the present study, therefore, will be concerned exclusively with the problem of establishing a definition of confirmation which satisfies the more severe formal conditions represented by (8.1), (8.2), and (8.3) together.

The fulfilment of these requirements, which may be regarded as general laws of the logic of confirmation, is of course only a necessary, not a sufficient, condition for the adequacy of any proposed definition of confirmation. Thus, *e.g.*, if "B confirms

[1] This was pointed out to me by Dr. Nelson Goodman. The definition later to be outlined in this essay, which satisfies conditions (8.1), (8.2) and (8.3), lends itself, however, to certain generalizations which satisfy only the more liberal conditions of adequacy just considered.

H " were defined as meaning " B logically entails H ", then the above three conditions would clearly be satisfied; but the definition would not be adequate because confirmation has to be a more comprehensive relation than entailment (the latter might be referred to as the special case of *conclusive* confirmation). Thus, a definition of confirmation, to be acceptable, also has to be materially adequate: it has to provide a reasonably close approximation to that conception of confirmation which is implicit in scientific procedure and methodological discussion. That conception is vague and to some extent quite unclear, as I have tried to show in earlier parts of this paper; therefore, it would be too much to expect full agreement as to the material adequacy of a proposed definition of confirmation; on the other hand, there will be rather general agreement on certain points; thus, *e.g.*, the identification of confirmation with entailment, or the Nicod criterion of confirmation as analyzed above, or any definition of confirmation by reference to a " sense of evidence ", will probably now be admitted not to be adequate approximations to that concept of confirmation which is relevant for the logic of science.

On the other hand, the soundness of the logical analysis (which, in a clear sense, always involves a logical reconstruction) of a theoretical concept cannot be gauged simply by our feelings of satisfaction at a certain proposed analysis; and if there are, say, two alternative proposals for defining a term on the basis of a logical analysis, and if both appear to come fairly close to the intended meaning, then the choice has to be made largely by reference to such features as the logical properties of the two reconstructions, and the comprehensiveness and simplicity of the theories to which they lead.

9. *The Satisfaction Criterion of Confirmation.*—As has been mentioned before, a precise definition of confirmation requires reference to some definite " language of science ", in which all observation reports and all hypotheses under consideration are assumed to be formulated, and whose logical structure is supposed to be precisely determined. The more complex this language, and the richer its logical means of expression, the more difficult it will be, as a rule, to establish an adequate definition of confirmation for it. However, the problem has been solved at least for certain cases: With respect to languages of a comparatively simple logical structure, it has been possible to construct an explicit definition of confirmation which satisfies all of the above logical requirements, and which appears to be intuitively rather adequate. An exposition of the technical details of this

definition has been published elsewhere;[1] in the present study, which is concerned with the general logical and methodological aspects of the problem of confirmation rather than with technical details, it will be attempted to characterize the definition of confirmation thus obtained as clearly as possible with a minimum of technicalities.

Consider the simple case of the hypothesis H: '(x)(Raven(x) $\supset$ Black(x))', where 'Raven' and 'Black' are supposed to be terms of our observational vocabulary. Let B be an observation report to the effect that Raven(a) . Black(a) . $\sim$ Raven(c) . Black(c) . $\sim$ Raven(d) . $\sim$ Black(d). Then B may be said to confirm H in the following sense: There are three objects altogether mentioned in B, namely a, c, and d; and as far as these are concerned, B informs us that all those which are ravens (*i.e.* just the object a) are also black.[2] In other words, from the information contained in B we can infer that the hypothesis H does hold true within the finite class of those objects which are mentioned in B.

Let us apply the same consideration to a hypothesis of a logically more complex structure. Let H be the hypothesis "Everybody likes somebody"; in symbols: '$(x)(Ey)$Likes(x, y)',

[1] In my article referred to in n. 1, p. 1. The logical structure of the languages to which the definition in question is applicable is that of the lower functional calculus with individual constants, and with predicate constants of any degree. All sentences of the language are assumed to be formed exclusively by means of predicate constants, individual constants, individual variables, universal and existential quantifiers for individual variables, and the connective symbols of denial, conjunction, alternation, and implication. The use of predicate variables or of the identity sign is not permitted.

As to the predicate constants, they are all assumed to belong to the observational vocabulary, *i.e.* to denote a property or a relation observable by means of the accepted techniques. ("Abstract" predicate terms are supposed to be defined in terms of those of the observational vocabulary and then actually to be replaced by their *definientia*, so that they never occur explicitly.)

As a consequence of these stipulations, an observation report can be characterized simply as a conjunction of sentences of the kind illustrated by '$P(a)$', '$\sim P(b)$', '$R(c, d)$', '$\sim R(e, f)$', etc., where 'P', 'R', etc., belong to the observational vocabulary, and 'a', 'b', 'c', 'd', 'e', 'f', etc., are individual names, denoting specific objects. It is also possible to define an observation report more liberally as any sentence containing no quantifiers, which means that besides conjunctions also alternations and implication sentences formed out of the above kind of components are included among the observation reports.

[2] I am indebted to Dr. Nelson Goodman for having suggested this idea; it initiated all those considerations which finally led to the definition to be outlined below.

i.e. for every (person) x, there exists at least one (not necessarily different person) y such that x likes y. (Here again, 'Likes' is supposed to be a relation-term which occurs in our observational vocabulary.) Suppose now that we are given an observation report B in which the names of two persons, say 'e' and 'f', occur. Under what conditions shall we say that B confirms H? The previous illustration suggests the answer: If from B we can infer that H is satisfied within the finite class $\{e, f\}$; *i.e.* that within $\{e, f\}$ everybody likes somebody. This in turn means that e likes e or f, and f likes e or f. Thus, B would be said to confirm H if B entailed the statement "e likes e or f, and f likes e or f". This latter statement will be called the development of H for the finite class $\{e, f\}$.—

The concept of *development of a hypothesis, H, for a finite class of individuals, C*, can be defined in a general fashion; the development of H for C states what H would assert if there existed exclusively those objects which are elements of C.—Thus, *e.g.*, the development of the hypothesis $H_1 =$ '$(x)(P(x) \vee Q(x))$' (*i.e.* "Every object has the property P or the property Q") for the class $\{a, b\}$ is '$(P(a) \vee Q(a)) \,.\, (P(b) \vee Q(b))$' (*i.e.* "$a$ has the property P or the property Q, and b has the property P or the property Q"); the development of the existential hypothesis H_2 that at least one object has the property P, *i.e.* '$(Ex)P(x)$', for $\{a, b\}$ is '$P(a) \vee P(b)$'; the development of a hypothesis which contains no quantifiers, such as H_3: '$P(c) \vee Q(c)$' is defined as that hypothesis itself, no matter what the reference class of individuals is.

A more detailed formal analysis based on considerations of this type leads to the introduction of a general relation of confirmation in two steps; the first consists in defining a special relation of direct confirmation along the lines just indicated; the second step then defines the general relation of confirmation by reference to direct confirmation.

Omitting minor details, we may summarize the two definitions as follows:

(9.1 Df.) An observation report B directly confirms a hypothesis H if B entails the development of H for the class of those objects which are mentioned in B.

(9.2 Df.) An observation report B confirms a hypothesis H if H is entailed by a class of sentences each of which is directly confirmed by B.

The criterion expressed in these definitions might be called the satisfaction criterion of confirmation because its basic idea consists in construing a hypothesis as confirmed by a given

observation report if the hypothesis is satisfied in the finite class of those individuals which are mentioned in the report.—Let us now apply the two definitions to our last examples: The observation report B_1: '$P(a) \,.\, Q(b)$' directly confirms (and therefore also confirms) the hypothesis H_1, because it entails the development of H_1 for the class $\{a, b\}$, which was given above.—The hypothesis H_3 is not directly confirmed by B, because its development—*i.e.* H_3 itself—obviously is not entailed by B_1. However, H_3 is entailed by H_1, which is directly confirmed by B_1; hence, by virtue of (9.2), B_1 confirms H_3.

Similarly, it can readily be seen that B_1 directly confirms H_2.

Finally, to refer to the first illustration given in this section: The observation report '$\text{Raven}(a) \,.\, \text{Black}(a) \,.\, \sim \text{Raven}(c) \,.\, \sim \text{Black}(c) \,.\, \sim \text{Raven}(d) \,.\, \sim \text{Black}(d)$' confirms (even directly) the hypothesis '$(x)(\text{Raven}(x) \supset \text{Black}(x))$', for it entails the development of the latter for the class $\{a, c, d\}$, which can be written as follows: '$(\text{Raven}(a) \supset \text{Black}(a)) \,.\, (\text{Raven}(c) \supset \text{Black}(c)) \,.\, (\text{Raven}(d) \supset \text{Black}(d))$'.

It is now easy to define disconfirmation and neutrality:

(9.3 Df.) An observation report B disconfirms a hypothesis H if it confirms the denial of H.

(9.4 Df.) An observation report B is neutral with respect to a hypothesis H if B neither confirms nor disconfirms H.

By virtue of the criteria laid down in (9.2), (9.3), (9.4), every consistent observation report, B, divides all possible hypotheses into three mutually exclusive classes: those confirmed by B, those disconfirmed by B, and those with respect to which B is neutral.

The definition of confirmation here proposed can be shown to satisfy all the formal conditions of adequacy embodied in (8.1), (8.2), and (8.3) and their consequences; for the condition (8.2) this is easy to see; for the other conditions the proof is more complicated.[1]

[1] For these proofs, see the article referred to in n. 1, p. 1. I should like to take this opportunity to point out and to remedy a certain defect of the definition of confirmation which was developed in that article, and which has been outlined above: this defect was brought to my attention by a discussion with Dr. Olaf Helmer.

It will be agreed that an acceptable definition of confirmation should satisfy the following further condition which might well have been included among the logical standards of adequacy set up in section 8 above: (8.4). If B_1 and B_2 are logically equivalent observation reports and B_1 confirms (disconfirms, is neutral with respect to) a hypothesis H, then B_2, too, confirms (disconfirms, is neutral with respect to) H. This condition is indeed satisfied if observation reports are construed, as they have been in this article, as classes or conjunctions of observation sentences. As was indicated at the end of n. 1, p. 108, however, this restriction of observation

Furthermore, the application of the above definition of confirmation is not restricted to hypotheses of universal conditional form (as Nicod's criterion is, for example), nor to universal hypotheses in general; it applies, in fact, to any hypothesis which can be expressed by means of property and relation terms of the observational vocabulary of the given language, individual names, the customary connective symbols for 'not', 'and', 'or', 'if-then', and any number of universal and existential quantifiers.

Finally, as is suggested by the preceding illustrations as well as by the general considerations which underlie the establishment of the above definition, it seems that we have obtained a definition

reports to a conjunctive form is not essential; in fact, it has been adopted here only for greater convenience of exposition, and all the preceding results, including especially the definitions and theorems of the present section, remain applicable without change if observation reports are given the more liberal interpretation characterized at the end of n. 1, p. 108. (In this case, if 'P' and 'Q' belong to the observational vocabulary, such sentences as '$P(a) \vee Q(a)$', '$P(a) \vee \sim Q(b)$', etc., would qualify as observation reports.) This broader conception of observation reports was therefore adopted in the article referred to in n. 1, p. 1; but it has turned out that in this case, the definition of confirmation summarized above does not generally satisfy the requirement (8.4). Thus, *e.g.*, the observation reports, $B_1 =$ '$P(a)$' and $B_2 =$ '$P(a) . (Q(b) \vee \sim Q(b))$' are logically equivalent, but while B_1 confirms (and even directly confirms) the hypothesis $H_1 =$ '$(x)P(x)$', the second report does not do so, essentially because it does not entail '$P(a) . P(b)$', which is the development of H_1 for the class of those objects mentioned in B_2. This deficiency can be remedied as follows: The fact that B_2 fails to confirm H_1 is obviously due to the circumstance that B_2 contains the individual constant 'b', without asserting anything about b: The object b is mentioned only in an analytic component of B_2. The atomic constituent '$Q(b)$' will therefore be said to occur (twice) inessentially in B_2. Generally, an atomic constituent A of a molecular sentence S will be said to occur inessentially in S if by virtue of the rules of the sentential calculus S is equivalent to a molecular sentence in which A does not occur at all. Now an object will be said to be mentioned inessentially in an observation report if it is mentioned only in such components of that report as occur inessentially in it. The sentential calculus clearly provides mechanical procedures for deciding whether a given observation report mentions any object inessentially, and for establishing equivalent formulations of the same report in which no object is mentioned inessentially. Finally, let us say that an object is mentioned essentially in an observation report if it is mentioned, but not only mentioned inessentially, in that report. Now we replace 9.1 by the following definition:

(9.1*a*) An observation report B directly confirms a hypothesis H if B entails the development of H for the class of those objects which are mentioned essentially in B.

The concept of confirmation as defined by (9.1*a*) and (9.2) now satisfies (8.4) in addition to (8.1), (8.2), (8.3) even if observation reports are construed in the broader fashion characterized earlier in this footnote.

of confirmation which also is materially adequate in the sense of being a reasonable approximation to the intended meaning of confirmation.

A brief discussion of certain special cases of confirmation might serve to shed further light on this latter aspect of our analysis.

10. *The Relative and the Absolute Concepts of Verification and Falsification.*—If an observation report entails a hypothesis H, then, by virtue of (8.1), it confirms H. This is in good agreement with the customary conception of confirming evidence; in fact, we have here an extreme case of confirmation, the case where B *conclusively confirms* H; this case is realized if, and only if, B entails H. We shall then also say that B *verifies* H. Thus, verification is a special case of confirmation; it is a logical relation between sentences; more specifically, it is simply the relation of entailment with its domain restricted to observation sentences.

Analogously, we shall say that B *conclusively disconfirms* H, or B *falsifies* H, if and only if B is incompatible with H; in this case, B entails the denial of H and therefore, by virtue of (8.1) and (9.3), confirms the denial of H and disconfirms H. Hence, falsification is a special case of disconfirmation; it is the logical relation of incompatibility between sentences, with its domain restricted to observation sentences.

Clearly, the concepts of *verification* and *falsification* as here defined are *relative;* a hypothesis can be said to be verified or falsified only with respect to some observation report; and a hypothesis may be verified by one observation report and may not be verified by another. There are, however, hypotheses which cannot be verified and others which cannot be falsified by any observation report. This will be shown presently. We shall say that a given *hypothesis is verifiable* (*falsifiable*) if it is possible to construct an observation report which verifies (falsifies) the hypothesis. Whether a hypothesis is verifiable, or falsifiable, in this sense depends exclusively on its logical form. Briefly, the following cases may be distinguished:

(*a*) If a hypothesis does not contain the quantifier terms "all" and "some" or their symbolic equivalents, then it is both verifiable and falsifiable. Thus, *e.g.*, the hypothesis "Object *a* turns blue or green" is entailed and thus verified by the report "Object *a* turns blue"; and the same hypothesis is incompatible with, and thus falsified by, the report "Object *a* turns neither blue nor green".

(*b*) A purely existential hypothesis (*i.e.* one which can be symbolized by a formula consisting of one or more existential quantifiers followed by a sentential function containing no

quantifiers) is verifiable, but not falsifiable, if—as is usually assumed—the universe of discourse contains an infinite number of objects.—Thus, *e.g.*, the hypothesis " There are blue roses " is verified by the observation report " Object *a* is a blue rose ", but no finite observation report can ever contradict and thus falsify the hypothesis.

(*c*) Conversely, a purely universal hypothesis (symbolized by a formula consisting of one or more universal quantifiers followed by a sentential function containing no quantifiers) is falsifiable but not verifiable for an infinite universe of discourse. Thus, *e.g.*, the hypothesis " $(x)(\text{Swan}(x) \supset \text{White}(x))$ " is completely falsified by the observation report $\{\text{Swan}(a), \sim \text{White}(a)\}$; but no finite observation report can entail and thus verify the hypothesis in question.

(*d*) Hypotheses which cannot be expressed by sentences of one of the three types mentioned so far, and which in this sense require both universal and existential quantifiers for their formulation, are as a rule neither verifiable nor falsifiable.[1] Thus, *e.g.*, the hypothesis " Every substance is soluble in some solvent "—symbolically ' $(x)(Ey)\text{Soluble}(x, y)$ '—is neither entailed by, nor incompatible with any observation report, no matter how many cases of solubility or non-solubility of particular substances in particular solvents the report may list. An analogous remark applies to the hypothesis " You can fool some of the people all of the time ", whose symbolic formulation ' $(Ex)(t)\text{Fl}(x,t)$ ' contains one existential and one universal quantifier. But of course, all of the hypotheses belonging to this fourth class are capable of being confirmed or disconfirmed by suitable observation reports; this was illustrated early in section 9 by reference to the hypothesis ' $(x)(Ey)\text{Likes}(x, y)$ '.

This rather detailed account of verification and falsification has been presented not only in the hope of further elucidating the meaning of confirmation and disconfirmation as defined above, but also in order to provide a basis for a sharp differentiation of two meanings of verification (and similarly of falsification) which have not always been clearly separated in recent discussions of the character of empirical knowledge. One of the two meanings of verification which we wish to distinguish here is the relative concept just explained; for greater clarity we shall sometimes

[1] A more precise study of the conditions of non-verifiability and non-falsifiability would involve technicalities which are unnecessary for the purposes of the present study. Not all hypotheses of the type described in (*d*) are neither verifiable nor falsifiable; thus, *e.g.*, the hypothesis ' $(x)(Ey)(P(x) \vee Q(y))$ ' is verified by the report ' $Q(a)$ ', and the hypothesis ' $(x)(Ey)(P(x) \,.\, Q(y))$ ' is falsified by ' $\sim P(a)$ '.

refer to it as *relative verification*. The other meaning is what may be called *absolute or definitive verification*. This latter concept of verification does not belong to formal logic, but rather to pragmatics [1]: it refers to the acceptance of hypotheses by "observers" or "scientists", etc., on the basis of relevant evidence. Generally speaking, we may distinguish three phases in the scientific test of a given hypothesis (which do not necessarily occur in the order in which they are listed here). The first phase consists in the performance of suitable experiments or observations and the ensuing acceptance of observation sentences, or of observation reports, stating the results obtained; the next phase consists in confronting the given hypothesis with the accepted observation reports, *i.e.* in ascertaining whether the latter constitute confirming, disconfirming or irrelevant evidence with respect to the hypothesis; the final phase consists either in accepting or rejecting the hypothesis on the strength of the confirming or disconfirming evidence constituted by the accepted observation reports, or in suspending judgment, awaiting the establishment of further relevant evidence.

The present study has been concerned almost exclusively with the second phase; as we have seen, this phase is of a purely logical character; the standards of evaluation here invoked—namely the criteria of confirmation, disconfirmation and neutrality—can be completely formulated in terms of concepts belonging to the field of pure logic.

The first phase, on the other hand, is of a pragmatic character; it involves no logical confrontation of sentences with other sentences. It consists in performing certain experiments or systematic observations and noting the results. The latter are expressed in sentences which have the form of observation reports, and their acceptance by the scientist is connected (by causal, not by logical relations) with experiences occurring in those tests. (Of course, a sentence which has the form of an observation report may in certain cases be accepted not on the basis of direct observation, but because it is confirmed by other observation reports which were previously established; but this process is illustrative of the second phase, which was discussed before. Here we are considering the case where a sentence is accepted directly "on the basis of experiential findings" rather than because it is supported by previously established statements.)

The third phase, too, can be construed as pragmatic, namely as consisting in a decision on the part of the scientist or a group of

[1] In the sense in which the term is used by Carnap in the work referred to in n. 1, p. 22.

scientists to accept (or reject, or leave in suspense, as the case may be) a given hypothesis after ascertaining what amount of confirming or of disconfirming evidence for the hypothesis is contained in the totality of the accepted observation sentences. However, it may well be attempted to give a reconstruction of this phase in purely logical terms. This would require the establishment of general "rules of acceptance"; roughly speaking, these rules would state how well a given hypothesis has to be confirmed by the accepted observation reports to be scientifically acceptable itself;[1] *i.e.* the rules would formulate criteria for the acceptance or rejection of a hypothesis by reference to the kind and amount of confirming or disconfirming evidence for it embodied in the totality of accepted observation reports; possibly, these criteria would also refer to such additional factors as the "simplicity" of the hypothesis in question, the manner in which it fits into the system of previously accepted theories, etc. It is at present an open question to what extent a satisfactory system of such rules can be formulated in purely logical terms.[2]

[1] A stimulating discussion of some aspects of what we have called rules of acceptance is contained in an article by Felix Kaufmann, 'The logical rules of scientific procedure', *Philosophy and Phenomenological Research*, June, 1942.

If an explicit definition of the degree of confirmation of a hypothesis were available, then it might be possible to formulate criteria of acceptance in terms of the degree to which the accepted observation reports confirm the hypothesis in question.

[2] The preceding division of the test of an empirical hypothesis into three phases of different character may prove useful for the clarification of the question whether or to what extent an empiricist conception of confirmation implies a "coherence theory of truth". This issue has recently been raised by Bertrand Russell, who, in ch. x of his *Inquiry into Meaning and Truth*, has levelled a number of objections against the views of Otto Neurath on this subject (*cf.* the articles mentioned in the next footnote), and against statements made by myself in articles published in *Analysis* in 1935 and 1936. I should like to add here a few, necessarily brief, comments on this issue.

(1) While, in the articles in *Analysis*, I argued in effect that the only possible interpretation of the phrase "Sentence S is true" is "S is highly confirmed by accepted observation reports", I should now reject this view. As the work of A. Tarski, R. Carnap, and others has shown, it is possible to define a semantical concept of truth which is not synonymous with that of strong confirmation, and which corresponds much more closely to what has customarily been referred to as truth, especially in logic, but also in other contexts. Thus, *e.g.*, if S is any empirical sentence, then either S or its denial is true in the semantical sense, but clearly it is possible that neither S nor its denial is highly confirmed by available evidence. To assert that a hypothesis is true is equivalent to asserting the hypothesis

At any rate, the acceptance of a hypothesis on the basis of a sufficient body of confirming evidence will as a rule be tentative, and will hold only "until further notice", *i.e.* with the proviso that if new and unfavourable evidence should turn up (in other words, if new observation reports should be accepted which disconfirm the hypothesis in question) the hypothesis will be abandoned again.

Are there any exceptions to this rule? Are there any empirical hypotheses which are capable of being established definitively, hypotheses such that we can be sure that once accepted on the basis of experiential evidence, they will never have to be revoked? Hypotheses of this kind will be called absolutely or definitively verifiable; and the concept of absolute or definitive falsifiability will be construed analogously.

While the existence of hypotheses which are relatively verifiable or relatively falsifiable is a simple logical fact, which was illustrated in the beginning of this section, the question of the existence of absolutely verifiable, or absolutely falsifiable, hypotheses is a highly controversial issue which has received a great deal of attention in recent empiricist writings.[1] As the problem

itself; therefore the truth of an empirical hypothesis can be ascertained only in the sense in which the hypothesis itself can be established: *i.e.* the hypothesis—and thereby *ipso facto* its truth—can be more or less well confirmed by empirical evidence; there is no other access to the question of the truth of a hypothesis.

In the light of these considerations, it seems advisable to me to reserve the term 'truth' for the semantical concept; I should now phrase the statements in the *Analysis* articles as dealing with confirmation. (For a brief and very illuminating survey of the distinctive characteristics of truth and confirmation, see R. Carnap, "Wahrheit und Bewährung," *Actes Ier Congrès Internat. de Philosophie Scientifique 1935*, vol. 4; Paris, 1936.)

(2) It is now clear also in what sense the test of a hypothesis is a matter of confronting sentences with sentences rather than with "facts", or a matter of the "coherence" of the hypothesis and the accepted basic sentences: All the logical aspects of scientific testing, *i.e.* all the criteria governing the second and third of the three phases distinguished above, are indeed concerned only with certain relationships between the hypotheses under test and certain other sentences (namely the accepted observation reports); no reference to extra-linguistic "facts" is needed. On the other hand, the first phase, the acceptance of certain basic sentences in connection with certain experiments or observations, involves, of course, extra-linguistic procedures; but this had been explicitly stated by the author in the articles referred to before. The claim that the views concerning truth and confirmation which are held by contemporary logical empiricism involve a coherence theory of truth is therefore mistaken.

[1] *Cf.* especially A. Ayer, *The Foundations of Empirical Knowledge* (New York, 1940); see also the same author's article, "Verification and Experience", *Proceedings of the Aristotelian Society* for 1937; R. Carnap,

is only loosely connected with the subject of this essay, we shall restrict ourselves here to a few general observations.

Let it be assumed that the language of science has the general structure characterized and presupposed in the previous discussions, especially in section 9. Then it is reasonable to expect that only such hypotheses can possibly be absolutely verifiable as are relatively verifiable by suitable observation reports; hypotheses of universal form, for example, which are not even capable of relative verification, certainly cannot be expected to be absolutely verifiable: In however many instances such a hypothesis may have been borne out by experiential findings, it is always possible that new evidence will be obtained which disconfirms the hypothesis. Let us, therefore, restrict our search for absolutely verifiable hypotheses to the class of those hypotheses which are relatively verifiable.

Suppose now that H is a hypothesis of this latter type, and that it is relatively verified, *i.e.* logically entailed, by an observation report B, and that the latter is accepted in science as an account of the outcome of some experiment or observation. Can we then say that H is absolutely confirmed, that it will never be revoked? Clearly, that depends on whether the report B has been accepted irrevocably, or whether it may conceivably suffer the fate of being disavowed later. Thus the question as to the existence of absolutely verifiable hypotheses leads back to the question of whether all, or at least some, observation reports become irrevocable parts of the system of science once they have been accepted in connection with certain observations or experiments. This question is not simply one of fact; it cannot adequately be answered by a descriptive account of the research behaviour of scientists. Here, as in all other cases of logical analysis of science, the problem calls for a "rational reconstruction" of scientific procedure, *i.e.* for the construction of a consistent and comprehensive theoretical model of scientific inquiry, which is then to serve as a system of reference, or a standard, in the examination of any particular scientific research. The

"Ueber Protokollsätze", *Erkenntnis*, vol. 3 (1932), and § 82 of the same author's *The Logical Syntax of Language* (see n. 1, p. 3). O. Neurath, "Protokollsätze", *Erkenntnis*, vol. 3 (1932); "Radikaler Physikalismus und 'wirkliche Welt'", *Erkenntnis*, vol. 4 (1934); "Pseudorationalismus der Falsifikation", *Erkenntnis*, vol. 5 (1935). K. Popper, *Logik der Forschung* (see n. 1, p. 4). H. Reichenbach, *Experience and Prediction* (Chicago, 1938), ch. iii. Bertrand Russell, *An Inquiry into Meaning and Truth* (New York, 1940), especially chs. x and xi. M. Schlick, "Ueber das Fundament der Erkenntnis", *Erkenntnis*, vol. 4 (1934).

construction of the theoretical model has, of course, to be oriented by the characteristics of actual scientific procedure, but it is not determined by the latter in the sense in which a descriptive account of some scientific study would be. Indeed, it is generally agreed that scientists sometimes infringe the standards of sound scientific procedure; besides, for the sake of theoretical comprehensiveness and systematization, the abstract model will have to contain certain idealized elements which cannot possibly be determined in detail by a study of how scientists actually work. This is true especially of observation reports: A study of the way in which laboratory reports, or descriptions of other types of observational findings, are formulated in the practice of scientific research is of interest for the choice of assumptions concerning the form and the status of observation sentences in the model of a "language of science"; but clearly, such a study cannot completely determine what form observation sentences are to have in the theoretical model, nor whether they are to be considered as irrevocable once they are accepted.

Perhaps an analogy may further elucidate this view concerning the character of logical analysis: Suppose that we observe two persons whose language we do not understand playing a game on some kind of chess board; and suppose that we want to "reconstruct" the rules of the game. A mere descriptive account of the playing-behaviour of the individuals will not suffice to do this; indeed, we should not even necessarily reject a theoretical reconstruction of the game which did not always characterize accurately the actual moves of the players: we should allow for the possibility of occasional violations of the rules. Our reconstruction would rather be guided by the objective of obtaining a consistent and comprehensive system of rules which are as simple as possible, and to which the observed playing behaviour conforms at least to a large extent. In terms of the standard thus obtained, we may then describe and critically analyze any concrete performance of the game.

The parallel is obvious; and it appears to be clear, too, that in both cases the decision about various features of the theoretical model will have the character of a convention, which is influenced by considerations of simplicity, consistency, and comprehensiveness, and not only by a study of the actual procedure of scientists at work.[1]

[1] A clear account of the sense in which the results of logical analysis represent conventions can be found in §§ 9-11 and 25-30 of K. Popper's *Logik der Forschung*. An illustration of the considerations influencing the

This remark applies in particular to the specific question under consideration, namely whether "there are" in science any irrevocably accepted observation reports (all of whose consequences would then be absolutely verified empirical hypotheses). The situation becomes clearer when we put the question into this form: Shall we allow, in our rational reconstruction of science, for the possibility that certain observation reports may be accepted as irrevocable, or shall the acceptance of all observation reports be subject to the "until further notice" clause? In comparing the merits of the alternative stipulations, we should have to investigate the extent to which each of them is capable of elucidating the structure of scientific inquiry in terms of a simple, consistent theory. We do not propose to enter into a discussion of this question here except for mentioning that various considerations militate in favour of the convention that no observation report is to be accepted definitively and irrevocably.[1] If this alternative is chosen, then not even those hypotheses which are entailed by accepted observation reports are absolutely verified, nor are those hypotheses which are found incompatible with accepted observation reports thereby absolutely falsified: in fact, in this case, no hypothesis whatsoever would be absolutely verifiable or absolutely falsifiable. If, on the other hand, some—or even all—observation sentences are declared irrevocable once they have been accepted, then those hypotheses entailed by or incompatible with irrevocable observation sentences will be absolutely verified, or absolutely falsified, respectively.

It should now be clear that the concepts of absolute and of relative verifiability (and falsifiability) are of an entirely different character. Failure to distinguish them has caused considerable misunderstanding in recent discussions on the nature of scientific knowledge. Thus, *e.g.*, K. Popper's proposal to admit as scientific hypotheses exclusively sentences which are (relatively) falsifiable by suitable observation reports has been criticized by means of arguments which, in effect, support the claim that scientific hypotheses should not be construed as being absolutely falsifiable—a point that Popper had not denied.—As can be seen from our earlier discussion of relative falsifiability, however, Popper's proposal to limit scientific hypotheses to the form of (relatively) falsifiable sentences involves a very severe restriction

determination of various features of the theoretical model is provided by the discussion in n. 1, p. 24.

[1] *Cf.* especially the publications by Carnap, Neurath, and Popper mentioned in n. 1, p. 116; also Reichenbach, *loc. cit.*, ch. ii, § 9.

of the possible forms of scientific hypotheses [1]; in particular, it rules out all purely existential hypotheses as well as most hypotheses whose formulation requires both universal and existential quantification; and it may be criticized on this account; for in terms of this theoretical reconstruction of science it seems difficult or altogether impossible to give an adequate account of the status and function of the more complex scientific hypotheses and theories.—

With these remarks let us conclude our study of the logic of confirmation. What has been said above about the nature of the logical analysis of science in general, applies to the present analysis of confirmation in particular: It is a specific proposal for a systematic and comprehensive logical reconstruction of a concept which is basic for the methodology of empirical science as well as for the problem area customarily called "epistemology". The need for a theoretical clarification of that concept was evidenced by the fact that no general theoretical account of confirmation has been available so far, and that certain widely accepted conceptions of confirmation involve difficulties so serious that it might be doubted whether a satisfactory theory of the concept is at all attainable.

It was found, however, that the problem can be solved: A general definition of confirmation, couched in purely logical terms, was developed for scientific languages of a specified and relatively simple logical character. The logical model thus obtained appeared to be satisfactory in the sense of the formal and material standards of adequacy that had been set up previously.

I have tried to state the essential features of the proposed analysis and reconstruction of confirmation as explicitly as possible in the hope of stimulating a critical discussion and of facilitating further inquiries into the various issues pertinent to this problem area. Among the open questions which seem to deserve careful consideration, I should like to mention the exploration of concepts of confirmation which fail to satisfy the general consistency condition; the extension of the definition of confirmation to the case where even observation sentences containing quantifiers are permitted; and finally the development of

[1] This was pointed out by R. Carnap; *cf.* his review of Popper's book in *Erkenntnis*, vol. 5 (1935), and "Testability and Meaning" (see n. 1, p. 5) §§ 25, 26. For the discussion of Popper's falsifiability criterion, see for example H. Reichenbach, "Ueber Induktion and Wahrscheinlichkeit", *Erkenntnis*, vol. 5 (1935); O. Neurath, "Pseudorationalismus der Falsifikation", *Erkenntnis*, vol. 5 (1935).

a definition of confirmation for languages of a more complex logical structure than that incorporated in our model.[1] Languages of this kind would provide a greater variety of means of expression and would thus come closer to the high logical complexity of the language of empirical science.

[1] The languages to which our definition is applicable have the structure of the lower functional calculus without identity sign (*cf.* n. 1, p. 108); it would be highly desirable so to broaden the general theory of confirmation as to make it applicable to the lower functional calculus with identity sign, or even to the higher functional calculus; for it seems hardly possible to give a precise formulation of more complex scientific theories without the logical means of expression provided by the higher functional calculus.

II.—THE PROBLEM OF UNREASONED BELIEFS (II.).

By W. T. Stace.

V

There is one rather obvious consideration which seems to contain some possibilities of helping us in our problem. Non-logical trains of thinking are usually, or at least frequently, guided by the principle of the association of ideas. If in a train of thought the successive steps do not lead to one another by links of logical connection, they may do so by links of association. Now the principle of the association of ideas is closely connected with the principle of induction. For if in our experience we find one thing x frequently accompanied by another thing y, then y becomes associated with x, and the thought of x tends to lead to the thought of y. But it is also true that, if x is frequently accompanied by y in our experience, there is a considerable probability that this is because there is a causal connection between them. In that case, if we *had* inductively argued from x to y, instead of merely passing from one to the other by association, we should have argued rightly. In this way non-logical thinking, guided only by association, may often lead us to correct conclusions.

It is clear that, if we are guided only by uncritically accepted associations, we shall make frequent mistakes. But it is also clear that such thinking will lead us to true conclusions *more often than can be explained by chance coincidence* . . . because we tend to connect by association things which are really connected in the outer world. Hence our thinking will have a certain degree of validity in the sense in which validity has been defined in this paper. The degree of validity may be low, owing to the frequency of mistakes, but it will be real.

I have here spoken as if associative thinking and inductive reasoning were two distinct kinds of thought which merely have a tendency to run parallel. As a matter of fact the connection is much closer than this. It seems to me that induction is in fact nothing more than a highly developed and critically safeguarded associative thinking. Association of the kind here envisaged is a very uncritical, undisciplined form of induction. Or induction is a disciplined form of association, in which we

have applied various rules designed to safeguard us against mistaking chance connections of events for causal connections. The canons of induction are formalized methods of discovering which of our associations lead to true propositions and which do not.

We can, if we wish, translate these remarks about associations of ideas into the more fashionable language of conditioned reflexes. The theory of conditioned reflexes is the behaviouristic version of the older theory of the association of ideas. The dog in Pavlov's famous experiment has been conditioned to react to the sound of the bell in the same manner as he would react, without conditioning, to the sight and smell of food. In the language of the older psychology he has associated the sound of the bell with the food. He may even be said—at least by a figure of speech—to have performed a crude induction. He has inductively concluded that bell-sounds are among the causes of dinner, or that they are regularly followed by dinner. Nor is he wholly mistaken. For there does exist a very remote and round-about chain of causal links, some of which are events in the mind or brain of the experimenter, between the bell-sound and the dinner. But he *is* mistaken in so far as he has concluded that there is a *direct* causal connection between the two, so that he expects that the bell-sound will *always* be followed by dinner. He is ignorant of the complexity of causes and of what Mill called counteracting causes. Hence he will be gravely disappointed when the bell sounds and no dinner appears. Nevertheless the dog will, during the period of experimentation at least, be right more often than can be explained by chance coincidence. After the experiments are over the disconnection of the bell-sounds and the dinner in his experience will tend to make the conditioned reflex disappear and thus to correct his mistake.

If an association is formed between x and y as a result of a chance coincidence between them—for instance, I may have a violent toothache in a certain room, and when I see the room again I may remember the toothache—then this association tends to be obliterated by subsequent appearances of x unaccompanied by y. Thus the associations which lead to true conclusions tend to remain fixed in the mind and the false associations to wither. Thereby the tendency of associative thinking to lead to true conclusions is enhanced, and their degree of validity increased.

As is well known, it is by mechanisms of this kind—whether we call them associations or conditioned reflexes—that animals learn. If we are allowed, at least by a figure of speech, to say that animals may correctly "believe" that fire will burn, that

certain foods which they see or smell will taste delicious, that certain other animals which they recognize on sight will devour them if they get the chance—if we are allowed, by a figure of speech, to say that the dog "believes" that his master will feed him and be kind to him and that schoolboys will throw stones at him—then all these beliefs will be unreasoned beliefs within the meaning of our definition. And their validity, the degree of which is apt to be, on the whole, rather high, will be easily explained by the principle of conditioned reflexes or association of ideas. And these considerations are helpful for our problem. For the human mind has evolved out of the animal mind. And many human unreasoned beliefs are very ancient and primitive, and their history may be supposed to go back almost to the animal stage. And it is not unreasonable to suppose that what in animals can only by a figure of speech be called "beliefs" are the rudimentary origins out of which have evolved some of the real unreasoned beliefs of human beings.

These considerations will at least partially explain the validity of *some* of our unreasoned beliefs. There is, for example, a class of beliefs which I have so far not mentioned and which may with some plausibility be called unreasoned. These are the beliefs which are embodied in judgments of perception. I see a box and judge that it is a box. It has been constantly insisted by philosophers that such judgments are not made as the result of a reasoning process. We are not, in making them, conscious of inferring anything. I believe that those philosophers who insist that they are not reasoned beliefs are only partly right. But let us for the moment assume that they are wholly right. Then these beliefs are not reached by reason. But neither are they cases of immediate knowledge. For what I immediately know are only certain directly presented sense-data, whereas the knowledge that "this is a box" involves various beliefs about facts which are not immediately presented, such as that the thing before me has a back and an inside, that it will feel hard if I touch it, etc. Hence, if we accept the view that the belief has not been reached by a reasoning process, it is then by definition an unreasoned belief. How has it been reached? If not by reason, then perhaps by association.[1] I have constantly found that visual appearances which are, in various specifiable ways, like this one, are followed by other experiences which are called seeing the back of the box when it is turned round, seeing its

[1] I speak tentatively here as if associative thinking and inductive reasoning could be sharply distinguished, which they cannot. I trust the matter is placed in a more correct light in section VI.

inside when it is opened, feeling it hard when it is touched, etc. The first visual appearance has become associated with the other experiences. I come to expect the latter because they are associated with the former.

But I do not think that this principle will explain the validity of *all* unreasoned beliefs. And I will now give some reasons for this opinion. In the first place, it will obviously only explain the validity of those beliefs which, although they have *not* been reached by induction, yet *could* have been reached by induction. Association only leads to true conclusions in those cases in which the associated phenomena are actually connected in nature. And in all such cases, of course, the true conclusions could have been reached by induction. Now it seems to me that there do exist unreasoned beliefs which never could have been reached by any induction. And if this is so, then their validity cannot be explained as due to associative thinking.

I think that there is at least one unreasoned belief which could never have been reached by induction, namely the belief that material objects exist during interperceptual intervals. Whether this belief can or cannot be reached by induction is a highly controversial question. It depends on what the belief is supposed to *mean*. If the proposition "M (a material object) exists when it is not being perceived" is wholly reducible, as most positivists contend that it is, to a series of hypothetical propositions such as "*if* I were in a certain place, then I *should* perceive M", then it can be proved by induction. I should myself contend, however, that this proposition is not wholly reducible to hypothetical propositions, but that it has, in addition, a categorical core which is irreducible to hypothetical propositions, *i.e.*, it asserts that something actually *is* the case, and not merely that *if* certain things were the case, *then* certain other things *would* be the case; in other words it alleges an actuality, not a mere set of possibilities. If this view is right, then in that case the belief that material objects exist when they are not being perceived cannot ever be proved by induction or in any other way. It is for this reason, because the categorical core is wholly unverifiable, that positivists assert that the proposition has no meaning except the series of hypothetical propositions. I disagree with this because I do not accept the positivist theory of meaning.[1] It will be seen that the question is very controversial, and I cannot argue it in this paper. But *if* my view is right, then the belief that material objects exist when they are

[1] See the article on Positivism previously referred to.

not perceived cannot be reached by induction. And it will also follow that its validity (assuming it to be valid) cannot be explained by associative thinking.

Another reason for thinking that, although the principle of association will explain the validity of some unreasoned beliefs, it will not explain the validity of them all, and cannot therefore provide a complete solution of our problem, is that many unreasoned beliefs seem to me to possess a very much higher degree of validity than can be explained by this principle. It is, of course, very difficult to prove this. I cannot claim that it is much more than a personal impression—in fact an unreasoned belief of my own. But it seems clear that the conclusions of the mathematical infant prodigy, which may be right ninety-nine times out of a hundred, have a far higher degree of validity than can be explained by association. Moreover, in this case, his results could not, so far as I know, be reached by induction. If they were obtained by reasoning at all it would presumably be by deduction. Thus the method of explanation by association seems wholly unsuitable in this case. It also seems to me that even the "flairs" and "intuitions" of the business man about economic situations or about human characters are something better than mere undirected associative thinking. And personally I should think that many moral "intuitions" have a higher degree of validity than can be accounted for by association. These remarks amount merely to the suggestion that some human unreasoned beliefs are on a higher level of validity than mere animal "beliefs" which are the results of random conditioning. If this is correct, it follows that we shall have to look for some additional principle of explanation, besides that of associative thinking.

Before beginning to look for such a principle, however, perhaps I had better clear up a possible misunderstanding of what has already been said. It may be said that I have explained the validity of some unreasoned beliefs by assimilating them to crude inductions, or by showing how they may have the same validity as crude inductions. And it may be said that this explains nothing since the validity of induction itself is not explained.

This is not a valid objection. I have purposely excluded the problem of induction from this paper. Moreover, if a man holds a belief for which we can see no grounds, and we ask him to justify it; and if he produces good evidence for it, *i.e.*, shows that it has the validity of a good induction, we consider that he has justified himself and explained how he came to a true conclusion. This is the procedure I am following in this paper.

Finally, the objection confuses two senses of the word validity. The problem of the validity of induction refers to validity in the logician's sense. It asks how induction is logically valid. But I am not asserting either that unreasoned beliefs are logically valid or that inductions are so. I am asserting that unreasoned beliefs are valid in the sense that they do actually lead to true conclusions more often than can be explained by chance coincidence. And therefore any question of their validity or non-validity in the logician's sense, or of the validity or non-validity of induction in this sense, is irrelevant to our problem.

VI

We may suppose that there is, in some psychological processes *a-b-c-d* . . . X, a degree of validity higher than can be explained by the supposition that the process is no more than a series of uncritically accepted associations or conditionings. This amounts only to supposing that some unreasoned human beliefs have more validity than those " beliefs " of animals which are reached by random conditioning. How is this higher degree of validity to be explained ? In this section I shall try to suggest a principle of explanation. And even if I am wrong in supposing that unreasoned beliefs have often a higher degree of validity than can be accounted for by associative thinking, yet I think that this second principle of explanation which I am about to suggest is one which does operate towards the production of unreasoned beliefs in addition to the principle of associative thinking.

We have already seen that no sharp line can be drawn between associative thinking and inductive reasoning. Associative thinking is a crude kind of induction, and what is ordinarily called induction is a refined and critically safeguarded kind of associative thinking. Thus the two kinds of thought-process shade off into one another. Now if it is the case that some unreasoned beliefs possess a higher degree of validity than can be explained by mere random associative thinking or conditioning, then my suggestion is that they have in fact been reached by something which approaches nearer the level of what is ordinarily called induction ; that the mere random conditioning which occurs even at the animal level is, in the case of human beings, supplemented by processes which in fact are reasoning processes. These reasoning processes, however, have a character which I shall call *unexpliciiness*, which causes them to be hidden from our view, so that we are not aware that they have taken place.

We then erroneously suppose that our beliefs are unreasoned. This amounts to saying that such unreasoned beliefs are not really unreasoned, but that they only appear to be so. And I think that, in certain cases, such as that of the mathematical prodigy, the results may be reached, not by unexplicit inductions, but by unexplicit deductions. I shall explain these suggestions in this section.

It seems to me that, in their consideration of actual thinking processes, philosophers have often been seriously misled by logicians—although this has perhaps been the fault of the philosophers rather than of the logicians. Logicians do not examine reasoning *processes*. A reasoning process is an event in time, something which happens in a man's mind or brain. It is the business of the psychologist, and not of the logician, to study it. Logicians are concerned with such things as propositions and the relation of implication between propositions. We will call all such things, which are the concern of the logician, "logical entities". Without committing ourselves to any view as to the status or nature of logical entities we may say that they are not identical with any psychological processes, or that, if they are, the logician is, at any rate, not concerned with the psychological aspect of them. Now I think that philosophers often do not clearly distinguish between logical entities and reasoning processes, and quite falsely suppose that the latter must have the same characteristics as the former; and that this may lead to serious mistakes.

It is characteristic of logical entities that they must be clearly differentiated into what I will call their constituents. For instance, the constituents of a chain of inferences are propositions —premisses and conclusions. The constituents of a proposition, again, are its subject, predicate, copula, etc. Now a proposition, for the logician, must be explicitly differentiated into subject, predicate, copula, etc. Anything which is not so differentiated is not a proposition. Similarly anything which is to be called an inference must be explicitly differentiated into premisses and conclusion. Thus Mr. W. E. Johnson says, "We may expect a storm when we notice the darkness of the sky, without at all having actually recalled past experiences of storms; in this case no inference is involved, since there has been only one assertion, namely what would constitute the conclusion without any other assertion that would constitute a premiss. In order to speak properly of inference in such cases, the minimum required is the assertion that the sky is cloudy and that *therefore* there will be a storm. Here we have two explicit assertions together with

the inference involved in the word 'therefore'".[1] (The word 'explicit' here should be noticed.)

Mr. Johnson is no doubt quite right in saying that we do not have an inference in the case mentioned if by inference he means, as I suppose he does, a logical entity. But the word inference is also used to mean a reasoning process. And if he means that in this case we do not have a reasoning process, then I think he is mistaken. Of course, it may be a mere verbal question whether the word inference is applicable in the case mentioned, although I think that common usage would support the view that it is. The man glances up at the clouds and says "it is going to rain". A very prosaic companion might say "you infer that from the darkness of the sky"? And the man would say "yes". But whatever word we use, the process of thought involved in the man's prediction is one which at any rate has the same validity as an inference and for the same logical reasons. His conclusion is just as valid when he does not consciously recall his past experience of storms as when he does. And his conclusion is reached *because of* the past experience and is validated *by* the past experience equally in both cases.

My point is that although it is an essential characteristic of all logical entities, such as propositions and inferences, to be explicitly differentiated into their constituents, yet actual human reasoning is quite different from this; and that we shall never understand how unreasoned beliefs can be valid if we do not realize this difference. It is true that, in certain relatively rare cases, human beings do reason by explicitly formulating and distinguishing in their minds the various propositions which constitute the chain of reasoning on which they are engaged. They do this if the subject-matter is abstruse, difficult, or unfamiliar. Philosophers, scientists, and other professional thinkers do it constantly. Ordinary men do it if faced with what may be for them especially tough problems. In such cases we consciously "think things out". But the vast majority of actual inferences in actual minds do not creep laboriously from premiss to premiss, and then, by means of a "therefore", to a conclusion. Premisses and conclusions are not clearly distinguished, much less the smaller constituents such as subjects and predicates. The whole process is apt to come in what seems like a single undivided act of thought.

The sheer rapidity of human thinking has something to do with this. A whole train of thought may flash through the mind at so great a speed that there is, as it were, no time to distinguish

[1] *Logic*, vol. 2, p. 4.

one constituent from another. Especially when the subject-matter is very familiar, or where similar inferences have frequently been made in the past and have become more or less habitual—like the prediction of the experienced seaman about the weather—the speed with which inferences are made is so great that their constituents cannot be individually noticed at all, and therefore cannot be distinguished from one another. The thought process is then undifferentiated into its constituents. We glance at the sky, and the impression of darkness and the expectation of rain come together into the mind as practically one thought. This is a very simple case which, when afterwards introspectively analyzed, may be recognized as complex but as containing only two constituents, a premiss and a conclusion. But I believe there are cases in which more or less long chains of reasoning, including ideas intermediate between the beginning premiss and the conclusion, may be gone through in the fraction of a second. In such cases we say "it came in a flash" or "I saw it all in a flash". In this last common phrase the words "it all" indicate that we afterwards realize the complexity of the thought-process which at the time seemed to come all in one piece.

I shall call all thought-processes in which the constituents are not clearly distinguished from one another, and which may be thought of therefore as undifferentiated, by the name "unexplicit thought-processes". Unexplicitness is a matter of degree. Thought-processes may be more or less differentiated or undifferentiated, more or less explicit or unexplicit. Also thought-processes of all kinds may be unexplicit. There may be unexplicit associative thinking, unexplicit induction, or unexplicit deduction.

To refer the unexplicitness of reasoning to the mere rapidity of human thinking would be to give a very superficial account of it. And I shall now refer to certain other characteristics which are of greater importance. I am compelled to describe these characteristics largely by means of metaphorical language. It is possible that if a behaviouristic version of what I am going to say in introspective terms could be given—as the language of unconditioned reflexes is the behaviouristic version of what may also be said in terms of the association of ideas—we might get something more satisfactory. I dare say this can, and someday will, be done.

In unexplicit thinking or reasoning, whether it is rapid or slow, the separate constituents of the process are either only very dimly distinguished or not distinguished at all. Consciousness shades off from a bright light, in which details stand out

separately, through a dimmer and dimmer twilight in which their separation grows more and more blurred, to an almost complete black-out in which the thought moves obscure, formless, and practically without parts.

There seem to be two opposite ways in which this may happen. First, a certain type of inference, originally explicitly performed, may by constant repetition have become habitual and automatic and so have sunk below the level of bright consciousness, and thus become unexplicit. In the opposite case a thought-process may be unexplicit because it has never risen above the level of consciousness at all. It exists only in the dim twilight where it has never been differentiated. It *can* be dragged up to the surface, and then the distinctness of its constituents will become for the first time visible, but this has not yet happened.

That reasoning processes may become unconscious and unexplicit through habit is not at all surprising. It has its analogues in the physical organism, which is what inclines me to think that it could be expounded in a purely behaviouristic version in which there would be no feeling of paradox. At present all accounts of unexplicit thought have an air of paradox because they seem to involve the contradiction of a more or less unconscious kind of consciousness. It seems paradoxical to speak of a thought-process having constituents which are yet not thought of or consciously distinguished. If we are not aware of thoughts, we are inclined to say, they cannot be thoughts at all. But if the same facts could be given in a behaviouristic version, from which the notion of consciousness has disappeared, we should avoid this trouble. Something of this sort has been done by Russell in the case of unconscious desire.[1]

I cannot give this behaviouristic version. But I can point to the fact that reasoning which has become unexplicit and largely unconscious through habit is analogous to certain well-known physical phenomena. When we learn to ride a bicycle, the different movements, pushing one leg down, letting the other come up, putting on the brake, are gone through in a conscious, laborious, step by step manner. Gradually these processes become automatic. We cease to notice what we are doing. We just "ride the bicycle", and this process, though it is really very complex, seems all one undivided act. Habit renders the differentiated consciousness of the separate acts involved unnecessary. We can also, when habit has taken over, perform the several acts with extreme rapidity. What I am here contending means little more than that reasoning activities are like

[1] *Analysis of Mind*, ch. iii.

bicycling activities. When the subject-matter on which we are engaged is difficult or unfamiliar, we reason slowly and laboriously in the logician's manner. The successive constituents of the reasoning then stand out in consciousness as differentiated judgments with differentiated terms, and as distinct passages from one judgment to another by means of a 'therefore'. But the reasoning may gradually become habitual and automatic, especially when the subject-matter is easy and familiar. We are then more or less unconscious of what we are thinking, and the terms, judgments, and 'therefores' telescope into what seems a single undivided act of thought.

Examples of this are mostly to be found in the region of every-day practical judgments. A common example is the inference we make, almost instantaneously, from the expression of a man's face, a frown, a scowl, a smile, to the belief that he is annoyed, pleased, scornful, etc. A swift judgment that a storm is due, based upon the darkness of the sky, may be another example. And I shall argue that ordinary judgments of perception, as when we recognize a tree or a table, involve inferences which are so completely habitual and automatic that all traces of explicit differentiation have disappeared from them. We then suppose the process of recognition to be immediate.

The opposite sort of unexplicitness, which is due to the fact that the thought-process has never yet risen above the level of consciousness may perhaps be made clearer by the following metaphor. Thoughts may be compared to organisms. They grow from undifferentiated foetus-like beginnings into highly differentiated forms. Under the influence of the logician's categories, philosophers are apt to believe that a judgment is either in the mind, complete, fully articulated into subject and predicate, or it is not in the mind at all. We are either thinking that S is P or we are not thinking it, they will say; and if we are thinking it, we must *first* think S and *then* think of P as its attribute. But this is quite a false picture of actual facts. It is a mere logical abstraction. The judgment which is clearly differentiated into subject, predicate, and copula, is only the last stage, the final and completely developed form of thinking. And the logician's proposition is an abstraction from this. Below this level are many grades of growth which become less and less clearly articulated, more and more blurred and formless, as we go down the scale; just as an embryo, though finally differentiated into arms, legs, eyes, etc., appears more and more formless and undifferentiated as we trace its development backwards towards its beginnings. At a low level of development a thought

will appear in consciousness in the form of a mere 'feeling'. This psychological fact is no doubt the basis of both Bradley's and Whitehead's doctrines of feeling. It is also perhaps the basis of the view of Spinoza that 'emotions' are confused thoughts. Now a vague, undeveloped, undifferentiated thought-process may be an embryonic reasoning process just as well as it may be an embryonic judgment or proposition. As a result of such an embryonic reasoning process a valid conclusion may emerge into full consciousness. In such a case we are not conscious of having gone through a process of reasoning at all, and the result will appear as an *unreasoned belief*.

Unexplicit reasonings are, I should say, always or generally unverbalized. This is another way in which they differ from what the logician calls inferences. It does not seem possible to conceive the logician's propositions and inferences as existing except explicitly in the form of symbols. For this reason some logicians tend to think of logic as concerned with sentences rather than with propositions. But whatever behaviourists may say, unexplicit reasonings take place without even "talking to oneself". Both the characteristics of rapidity and unexplicitness ensure this. The rapidity may be so great that it would be absolutely impossible to frame the successive phases in words, or even in the incipient motions of pronouncing words, or in separate images. It may be the case that bodily movements of some kind accompany every phase of thought,[1] which is the ground for the hope that everything here said could be given in a behaviouristic version. But the suggestion that the bodily movements which accompany thought are merely those of speaking is certainly much too crude. The unexplicitness of the reasonings as well as their mere rapidity ensures their being unverbalized. For the separate terms and judgments, being undivided from one another in thought, cannot find expression in a series of words divided from one another. Division into words presupposes division of the thought-content into its separate constituents, and where this is as yet absent there cannot be verbalization. As to the truth of all these statements, I can, of course, only appeal to the reader's experience. The experience of having a thought and of meeting the greatest difficulty in finding words for it—the thought thus clearly existing before the words—is surely common to all reflective persons, and especially to writers. This is in fact the great difficulty of the writer and the source of most of the heavy mental labour he has to undertake.

[1] As far as the argument of this paper is concerned it might be the case that the bodily movements *are* the thoughts, though I think myself that this is absurd.

It seems to me that these considerations help to clear up some of the problems about perception and judgments of perception. I see a chair or box and recognize it for what it is. Is this a case of immediate knowledge? Those philosophers who make what they call sense-data the basis of physical knowledge say that it is not. But there are other philosophers who are constantly protesting against this view. They insist that we do not see visual sense-data, but chairs and boxes. They are impressed by the apparent immediacy of sense-perception. No one, they say, on seeing his breakfast table in the morning, goes through any reasoning process in order to arrive at the conclusion that it is his breakfast table. The very suggestion of such a thing seems ridiculous. And we must not, as Professor Price has said, allow philosophers to bully us into admitting that we are thinking when we are not thinking.

In my view the difficulty is cleared up by the concept of unexplicit reasonings. Those who protest against the sense-datum theory are right because there is no conscious or explicit reasoning involved in perception. But those who insist that perception is not immediate knowledge but is mediate knowledge based upon the immediate apprehension of sense-data are also right, because unexplicit induction is involved. The seaman's judgment of the weather from the appearance of the dark clouds has from habit become more or less automatic and undifferentiated, and it is apt to seem *almost* immediate. Yet here we can faintly feel that there is a passage from the impression of the dark sky to the expectation of rain. But ordinary judgments of perception are judgments which all men make every moment of their lives from childhood to death. Habit has therefore made them completely automatic. The sense of a passage from one thought to another is entirely lost, wholly sunk below the level of consciousness. It has completely disappeared. The apprehension of the chair or the box seems to be immediate. Yet there *is* a passage, and this passage is of the nature of an induction. It is an unexplicit induction. In the previous section I suggested that, if judgments of perception are not reached by inference, they may be reached by unexplicit associative thinking. I am now suggesting that unexplicit inductions are involved. It will be remembered, however, that no sharp line can be drawn between associative thinking and inductive reasoning, but that the latter type of thinking is especially characteristic of the human animal.

There is no sharp line between explicit and unexplicit reasonings. It is a matter of degree. The passage in the case of sense-

perception is practically instantaneous. Yet it does involve a passage from a sense-datum to such expectations or "takings for granted" as that the object has a back, an inside, etc. Suppose now I see a chocolate box and expect that it will contain chocolates (in which, of course, I may be wrong). Are we to say that this is a case of perception or an inference? I think that most people would call it an inference. Yet there is no difference of principle between the two cases. In the one I pass from the immediately given appearance to the expectation of a back, an inside, etc. In the other I pass from the immediately given appearance to the expectation of chocolates. But the latter sort of inference is not nearly so frequent as that involved in ordinary judgments of perception, with the result that habit has not made it so instantaneous, automatic, unconscious, and undifferentiated. Hence we tend to call it an inference and not a perception. A third case, a little higher up in the scale of differentiation, might be that in which we notice a motor car passing in bright sunshine with rain drops on its wind-shield. We jump at once to the conclusion that it has been raining a few miles away. I think every one would call this an inference, and not a case of perception. And yet here again there is no difference of logical principle. And since we see that we cannot draw any sharp line between cases of perception and inferences, but that the one shades off into the other, on what principle are we to deny that perception itself involves and contains unexplicit inference?

What is a case of more or less laborious inference to one person may be a case of almost immediate perception or "intuition" to another. This is relative to the background of experience of the persons in question. An expert will recognize a botanical specimen at sight. He simply perceives it as what it is. His judgment is a judgment of perception. A layman may have to smell, touch, measure the specimen, count its petals, and then conclude that it is of such and such a species. In his case it is a relatively complicated inference. A young civil servant presented with an administrative problem, has to think out, step by step, what ought to be done. But an experienced administrator sees the solution at a glance.

It seems to me that most of the various "intuitions", "hunches", and "flairs" that we meet with in daily life are to be explained by similar considerations; and that either the principle of associative thinking, or that of unexplicit reasonings, or a combination of both, will account for the validity of the majority of our unreasoned beliefs. Even the performance of the mathematical prodigy can perhaps be set down as a case of

unexplicit deduction. There are, however, two species of unreasoned beliefs regarding which some further remarks must be made. These are moral intuitions, and what have been referred to as instinctive beliefs or common-sense beliefs about the external world.

VII

Moral intuitions seem to me to differ from other intuitions chiefly by the fact that they are the work of long generations of men rather than of single individuals. It should not take very long for men to learn from experience that the breaking of promises is a habit which usually leads to disagreeable consequences for a majority of the persons concerned. But the belief that the soul is more noble than the body is a vast vague generalization about life which may well have been the result of thousands of years of the cumulative experience of living. In principle it must have been reached by procedures which, however confused, tentative, fumbling, can be dimly recognized by us as inductive in their nature. But here the inductions can have been neither rapid nor automatic. Reflective or otherwise exceptional men must have noticed many facts about themselves and their fellow men which they epitomized in wise saws about human life. On the basis of such facts they must have framed such general, and no doubt hazy, ideas as that those who aim mostly at sensual pleasures do not as a rule really live happily. There are many observable facts about them which point to this. They are apt to be fretful, feverish, quickly dissatisfied, easily disillusioned, and when advancing years dull the senses, they have nothing left in life which yields satisfaction. Also their activities frequently tend to hurt others, and activities hurtful to others must very early have been recognized as generally disadvantageous to the group or society and therefore to be condemned. Those, on the other hand, who cultivate the things of the mind and spirit do no harm to others and find for themselves a more lasting kind of satisfaction. They live more quietly and are more contented. On the whole their lot, though less spectacular, is preferable. Their joys are more subtle and exquisite. And as Plato and Mill have pointed out, they have experienced both the pleasures of the mind and the body, whereas the sensualist knows only the latter; and knowing both they find the former more satisfying. These and many other reflections and observations of a similar kind, accumulating from generation to generation, must have built up the belief that spiritual things are in some way intrinsically finer than material things, or in other words that the soul is nobler than the body.

The 'inductions', of course, were not carried out in any scientific spirit. They were not guided by any specifically formulated canons. They were not safe-guarded by any collection of 'negative instances'. They were very rough and ready inductions by simple enumeration. But so have been most of the everyday inductions of practical men on which the ordinary conduct of life is based—such as that certain foods will not poison us, that certain herbs will cure certain diseases, that our friends will as a rule behave in certain predictable ways, that the spring will thaw the snows. The logic of them does not for that reason cease to be inductive.

These inductions about life, which have resulted in moral intuitions, cannot have been wholly automatic and unconscious. Rather they were slow and laborious. And therefore the fact that they are now apt to appear intuitive and immediate must be explained in some other way. Perhaps we may explain it by pointing out that the vast majority of the particular observations on which they were founded have long ago been forgotten and lost in the mists of time. In rude ages of the past wise men formulated these conclusions on the basis of their observations of life. They handed them on to their children and successors who had neither the wit nor the interest to understand the reasons for them. They were wise saws given out dictatorially by those who were supposed to know. They became traditions. And as with all traditions the grounds and reasons on which they were originally based came to be forgotten. Thus they appeared to men as mysterious and immediately known truths. Not only their grounds but their very meaning was forgotten; so that now the most learned philosophers do not know what they mean, and will deny that their meaning has anything to do with the consequences of actions. But gradually, as the cumulative experience of living piled up, these moral conclusions bit into the consciousness of men, became habitual thoughts. Men came to feel them in their bones. They came to seem intuitively known truths. And superstitious philosophers were found to invent mystic faculties of the mind to account for our knowledge of them, to hail them as *a priori* synthetic propositions, or as simple revelations from heaven.

VIII

Among what Russell called instinctive beliefs I shall discuss only two, and these very briefly—belief in other minds, and belief in the existence of material objects when they are not being perceived.

XUM

It seems to me that our belief in the minds of other persons *is* reached, in spite of the almost unanimous opinion of philosophers to the contrary, by means of a reasoning process, to wit, an analogy based upon the observation of bodily behaviour. But the process is, of course, wholly unexplicit, so that philosophers commonly regard it as instinctive or immediate. As to the general opinion of philosophers to the contrary, I have already accounted for it. Their tendency is to regard as immediate whatever they cannot explain by the most crude, obvious, and explicit kinds of reasoning. They imagine that all reasoning is like the logician's inferences. You either reason like a logician or you do not reason at all. They have no conception of the infinite subtlety, of the nuances and gradations of thinking, of the real human mind. Every piece of knowledge which is found in the mind they must needs force into one or the other of the only two pigeon-holes they have ever heard of, "immediate" and "reasoned", where "reasoned" means explicitly reasoned and preferably proceeding deliberately according to the elaborate canons of an Aristotle, a Mill, or a Russell. They forget that just as men can digest food without any knowledge of physiology or anatomy and without being aware of the processes going on in their bodies, so men can and have reasoned for hundreds of thousands of years without any knowledge of logic and without being aware of the processes going on in their minds. Consequently all the beliefs which men have but which have not been explicitly reasoned out—and this means the vast majority of human beliefs—are a mystery to them. There is, however, no more mystery about our knowledge of other minds than there is about our knowledge that "this is a chair". Both are cases of unexplicit inductions or analogies which, because they are instantaneous and unexplicit, appear to be unreasoned.

It may be objected that our belief in other minds comes in early infancy, and that we can hardly suppose that babes in the cradle reason. But in the first place I see no reason to doubt that infants reason, although they have not read Aristotle. They recognize objects very clearly, and this involves unexplicit reasoning or at least valid associative thinking, which is the same thing as crude inductive reasoning. They soon learn by experience that crying usually produces food. And learning by experience is either an inductive process or a process of associative thinking which cannot be distinguished from induction. In the second place it seems probable that human beings do not reach a belief in other minds till much later than is commonly supposed. They react appropriately to their parents and other persons at

a very early age, but at what age do they come to understand a truth so very abstruse and abstract as that "other people have minds"? It is comparatively late in life that a human being discovers his own mind.

Belief that material objects exist when they are not being perceived is, however, in a very different case. The question is a very complicated one which I cannot discuss at length here.[1] I shall only briefly indicate the way in which I think it ought to be treated. This belief cannot be explained either by the principle of associative thinking or by the principle of unexplicit induction. It falls therefore entirely outside any explanations given so far in this paper, and demands special treatment. For unless it is wholly analysable into a series of hypothetical propositions, which in my view it is not, it cannot be proved by any kind of induction or any sort of argument. It is not merely that it cannot be proved conclusively. The case is that there is not, and can never be, any evidence whatever which renders it even faintly probable or improbable (which is why the positivist says the categorical core of the belief has no meaning). And since it cannot be *proved* by explicit induction, it cannot be *reached* by unexplicit induction. Hence all the suggestions made so far in this paper altogether fail as regards this one unreasoned belief. And yet it is an unreasoned belief; for it is neither immediate nor reasoned.

There are, therefore, only two courses open to us. On the one hand we can set it down as a mysterious revelation of a supernatural faculty, we can call it *a priori;* or we can refuse to face the problem by appealing to 'common sense' (itself conceived as a mysterious source of revelation). Or on the other hand, we can say that this belief is groundless and has no validity, although it must have psychological causes. Taking the latter course, I should argue that the nature of the psychological causes will be of the kind commonly called 'pragmatic'. Not that it makes any difference to our 'practical' activities whether or not objects exist when they are not perceived. As long as the table is here when I come into the room and want to write on it, I do not care where it is, or whether it exists, when I am out of the room. The pragmatic grounds of the belief that it does exist when it is not perceived will probably have to do with considerations of 'simplicity'. It is simpler and easier to think of the world as continuously existing than to think of it as intermittent. A

[1] I have discussed it fully in two books, *The Nature of the World*, and (an earlier version) *The Theory of Knowledge and Existence*, without having, so far as I know, influenced a single reader in the direction of my opinion.

continuous straight line drawn on a piece of paper is less complex than a broken line, and therefore requires less mental effort to grasp. The eye and the imagination run smoothly along it without experiencing the succession of jolts entailed by the broken line. So it is with our picture of the world. And since it is equally open to us, so far as logic or evidence are concerned, to believe either in the continuous or the intermittent world, since we are not inclined one way or the other by even the slightest rational consideration of any kind, the comparatively minute psychological advantages of simplicity and economy of mental effort are sufficient to tilt the scale in favour of a continuous world. It is a matter of great interest to note that, in so far as psychological causes of this kind have actually produced our belief in a continuous world, the psychological processes involved must have been entirely unexplicit, unconscious, undifferentiated. This is an example of the principle, suggested earlier, that *any* process of thinking, and not merely a reasoning process, may be unexplicit. Of course no one would suggest that our savage ancestors *thought out*, or were conscious of, the pragmatic advantages involved in the choice of a continuous world. It is the unexplicitness of the thought-process which explains the apparently instinctive character of the belief.

The question of the validity of the process which leads to this unreasoned belief will depend upon whether the belief itself is true. And the question of its truth depends on the definition we give to the term truth. If truth means the correspondence of facts to ideas, then there is no reason for supposing that the belief is true or that the process has any validity. But if the statement that an idea is true is taken to mean only that it is convenient,—for example, that it simplifies our thinking, while not clashing with any observable fact,—then the belief may be called true, in this sense, and the process valid.

IX

The first of the problems stated in section 2 was: How is it that thought-processes which are, or at least *appear* to be, non-logical, can nevertheless be valid in the sense that they have a tendency to lead to truths? According to the tentative suggestions which I have put forward there are at least four different kinds of psychological process which have this tendency, *viz.*,

(1) Associative thinking or conditioning.
(2) Unexplicit inductive reasoning.
(3) Unexplicit deductive reasoning. This was suggested as a possible explanation of the performances of mathematical prodigies. It may

also account for the results of the more intuitive mathematicians, and for other unreasoned beliefs.

(4) Pragmatic thinking. This is, I am afraid, an ugly phrase. I mean by it any process of thought which leads men to hold a belief because the belief, whether true or not, produces advantages in thinking or acting or both. This seemed to me to explain our unreasoned belief in the existence of material objects when not being perceived. This was the only example I gave, but there may be many others.

These four kinds of thinking are, doubtless, not mutually exclusive. In actual thinking any or all of them can probably be combined. Thus no sharp distinction between (1) and (2) is possible, and a particular process of thought may sometimes be set down as either or both. Again, in an actual train of unexplicit thought, induction and deduction may of course be used in combination in the same way as they often are in explicit thinking. Nor is there any reason to suppose that pragmatic thinking must, when it occurs, be purely pragmatic. We may hold a belief both because there is evidence for it and because it is advantageous. In general, so far as I can see, *any* sort of pure or mixed psychological process which can occur explicitly can also occur unexplicitly. But these are, of course, ultimately questions for the psychologist.

It will be seen that, although we spoke at the beginning of this paper of "unreasoned beliefs", and of their being reached by "non-logical psychological processes", yet, if my suggestions are anywhere near the truth, neither of these expressions is strictly correct. Except perhaps in the case of beliefs produced by pure random conditioning, the so-called unreasoned beliefs only *appear* to be unreasoned, and the psychological processes only *appear* to be non-logical. Most of them are in their essence reasoning processes, though the reasoning may be more or less crude. We are not aware of their rational character, or even of their existence, because of their unexplicitness. In the rest of this article I shall however continue to use the expression "unreasoned beliefs" as a short and convenient label for the various kinds of beliefs which we have been discussing.

X

The second of our two problems was: What degrees of validity have the various kinds of psychological processes concerned? What relative weights ought we to attach to various unreasoned beliefs either in our philosophical constructions or in our practical lives?

As to practical living, it is unlikely that the philosopher, even if he has successfully solved the problem of unreasoned beliefs, will be able to offer advice of much value to the business man, the statesman, or the man in the street. They will rely on their intuitions to the extent that their experience and good sense (itself a psychological process having a tendency to lead to true beliefs) direct them. And they develop a greater or less practical skill in this matter. But even practical men might sometimes do better than they do if they guided their actions by certain conclusions which inevitably follow from the analysis we have given. For instance, if our analysis has been correct, or even roughly on the right lines, it will inevitably follow from it that *no belief based on an intuition can ever be as certain as the same belief would be if it were based on careful and conscious reasoning.* For the so-called intuition has in fact been reached by unexplicit reasoning, and its truth, if it is true, has been due to this fact. Therefore it would have been more reliable if it had been reached by the same reasoning explicitly set out and carefully checked and safe-guarded. If Hitler had been intelligent enough, and philosopher enough, to look at his own famous intuitions in this light, he would have trusted them less, and perhaps thought more, and the history of the world might have been different. And an understanding of this matter would save men from a superstitious worship of their own intuitions and from the fatal delusion—to which some of the most remarkable characters in history have been liable—that they are the special agents of God, and their convictions the immediate revelations of heaven.

But leaving the practical man aside, we may now consider the impact of our conclusions on questions of philosophical method. Nothing is more common than for philosophers to weave their unreasoned beliefs freely into their philosophical constructions and to claim for them any degree of validity which their fancy dictates. Thus Whitehead claims that his, or our, moral, æsthetic, and religious intuitions are to be considered as evidence for conclusions about the nature of the universe on a par with, or superior to, the evidence of the senses. Whitehead is only making explicit the common procedure of idealistic and 'tender-minded' philosophers since the world began. The 'tough-minded' philosophers do the same thing, the only difference being that, as we should expect, their intuitions are themselves more tough-minded; and they usually avoid the word intuition, because they think it sounds tender-minded. Their unreasoned beliefs may be carefully camouflaged and hidden from sight, only appearing in the guise of some unexpressed *bias*—in favour of

materialism, or realism, or against 'metaphysical' beliefs. Or they may be frankly avowed, as with Russell's 'instinctive beliefs' or Moore's 'common sense'. The claims made by philosophers on behalf of common sense are perhaps the most doubtful and ambiguous of all. There are scarcely any philosophers who do not in effect claim that their conclusions are supported by common sense, notwithstanding that all these conclusions differ from one another. Whitehead, just as much as Moore, claims that his metaphysics agrees at critical points with common sense, as well as with moral, æsthetic, and religious intuitions. The only difference between the so-called instinctive beliefs and these other intuitions is that the deliverances of the former concern lowly and humble matters, such as the non-mental character of chairs and tables, whereas the deliverances of the more pretentious intuitions tell us about the ultimate nature of God and value. And instinctive beliefs are apt to pronounce in favour of realism, while the more pretentious intuitions are likely to favour some kind of idealism. Whitehead's moral, æsthetic, and religious intuitions make him a panpsychist, while his common sense makes him a realist.

Probably it is inevitable that philosophers, as well as practical men, should rely on their unreasoned beliefs. The question, however, is: Can we, as a result of our analysis of the nature of unreasoned beliefs, give any indication of the *degrees* of validity which attach to this or that particular kind of unreasoned belief? Can we say anything which could put a stop to the present practice of treating them all alike as final authorities, without any discrimination of their respective merits, so that philosophy is simply a battleground of final authorities which contradict one another? Such a programme, if successfully carried out, might well revolutionize philosophy, turn it into a respectable body of knowledge instead of a brawl of contending opinions. Most certainly such a programme cannot be implemented here and now. I suggest it as a field for fruitful enquiry. And I emphasize again the hesitant and tentative character of the following points, which are no more than preliminary suggestions.

(1) No philosopher is entitled to claim, as Whitehead seems to do, that his unreasoned beliefs, whether he call them intuitions, common sense beliefs, instinctive beliefs, or anything else, are to be taken as *basic* evidence for anything. For whatever their value may be, they are not basic, *i.e.*, they are not cases of immediate knowledge which can be taken as certain, so that chains of reasoning can start *from* them. For they are themselves the results of reasoning processes which presumably start from evidence which is more basic.

As to practical living, it is unlikely that the philosopher, even if he has successfully solved the problem of unreasoned beliefs, will be able to offer advice of much value to the business man, the statesman, or the man in the street. They will rely on their intuitions to the extent that their experience and good sense (itself a psychological process having a tendency to lead to true beliefs) direct them. And they develop a greater or less practical skill in this matter. But even practical men might sometimes do better than they do if they guided their actions by certain conclusions which inevitably follow from the analysis we have given. For instance, if our analysis has been correct, or even roughly on the right lines, it will inevitably follow from it that *no belief based on an intuition can ever be as certain as the same belief would be if it were based on careful and conscious reasoning.* For the so-called intuition has in fact been reached by unexplicit reasoning, and its truth, if it is true, has been due to this fact. Therefore it would have been more reliable if it had been reached by the same reasoning explicitly set out and carefully checked and safe-guarded. If Hitler had been intelligent enough, and philosopher enough, to look at his own famous intuitions in this light, he would have trusted them less, and perhaps thought more, and the history of the world might have been different. And an understanding of this matter would save men from a superstitious worship of their own intuitions and from the fatal delusion—to which some of the most remarkable characters in history have been liable—that they are the special agents of God, and their convictions the immediate revelations of heaven.

But leaving the practical man aside, we may now consider the impact of our conclusions on questions of philosophical method. Nothing is more common than for philosophers to weave their unreasoned beliefs freely into their philosophical constructions and to claim for them any degree of validity which their fancy dictates. Thus Whitehead claims that his, or our, moral, æsthetic, and religious intuitions are to be considered as evidence for conclusions about the nature of the universe on a par with, or superior to, the evidence of the senses. Whitehead is only making explicit the common procedure of idealistic and 'tender-minded' philosophers since the world began. The 'tough-minded' philosophers do the same thing, the only difference being that, as we should expect, their intuitions are themselves more tough-minded; and they usually avoid the word intuition, because they think it sounds tender-minded. Their unreasoned beliefs may be carefully camouflaged and hidden from sight, only appearing in the guise of some unexpressed *bias*—in favour of

materialism, or realism, or against 'metaphysical' beliefs. Or they may be frankly avowed, as with Russell's 'instinctive beliefs' or Moore's 'common sense'. The claims made by philosophers on behalf of common sense are perhaps the most doubtful and ambiguous of all. There are scarcely any philosophers who do not in effect claim that their conclusions are supported by common sense, notwithstanding that all these conclusions differ from one another. Whitehead, just as much as Moore, claims that his metaphysics agrees at critical points with common sense, as well as with moral, æsthetic, and religious intuitions. The only difference between the so-called instinctive beliefs and these other intuitions is that the deliverances of the former concern lowly and humble matters, such as the non-mental character of chairs and tables, whereas the deliverances of the more pretentious intuitions tell us about the ultimate nature of God and value. And instinctive beliefs are apt to pronounce in favour of realism, while the more pretentious intuitions are likely to favour some kind of idealism. Whitehead's moral, æsthetic, and religious intuitions make him a panpsychist, while his common sense makes him a realist.

Probably it is inevitable that philosophers, as well as practical men, should rely on their unreasoned beliefs. The question, however, is: Can we, as a result of our analysis of the nature of unreasoned beliefs, give any indication of the *degrees* of validity which attach to this or that particular kind of unreasoned belief? Can we say anything which could put a stop to the present practice of treating them all alike as final authorities, without any discrimination of their respective merits, so that philosophy is simply a battleground of final authorities which contradict one another? Such a programme, if successfully carried out, might well revolutionize philosophy, turn it into a respectable body of knowledge instead of a brawl of contending opinions. Most certainly such a programme cannot be implemented here and now. I suggest it as a field for fruitful enquiry. And I emphasize again the hesitant and tentative character of the following points, which are no more than preliminary suggestions.

(1) No philosopher is entitled to claim, as Whitehead seems to do, that his unreasoned beliefs, whether he call them intuitions, common sense beliefs, instinctive beliefs, or anything else, are to be taken as *basic* evidence for anything. For whatever their value may be, they are not basic, *i.e.*, they are not cases of immediate knowledge which can be taken as certain, so that chains of reasoning can start *from* them. For they are themselves the results of reasoning processes which presumably start from evidence which is more basic.

(2) The proper procedure of philosophers is to state their intuitions, try to bring into the full light of consciousness the prior reasonings on which they were more or less unconsciously based, and then to test this reasoning by the ordinary canons of good logic or scientific method. This amounts to saying that they should attempt to rationally reconstruct their unreasoned beliefs. If they cannot successfully do so, their intuitions are suspect. They are not necessarily false, or even groundless, for they may have had grounds which the philosopher is now unable to drag up into the light of consciousness. But until he *can* do this, very little weight ought to be attached to his intuitions.

(3) No intuition can claim that it is as certain as a belief reached by careful and deliberate reasoning. In general, intuition is more fallible than careful reasoning. For intuitions are themselves based on unexplicit reasonings, and these are necessarily more or less rough and ready, crude, unsifted, unsafeguarded by the canons of good thinking.

(4) The evidence of the senses is as a rule more reliable than the evidence of moral æsthetic, and religious intuitions and, in case of a conflict, philosophers are not entitled to claim a superior rôle for the latter. For these intuitions are themselves as a rule based in the end on the evidence of the senses. It is not meant to deny the possibility that there may be *basic* evidence which is non-sensuous, for instance, introspective data or impressions of what Locke called reflection. But the danger of mistaking for non-sensuous data what are in fact unreasoned beliefs which are ultimately founded on sense-experience is greater than the danger of overlooking non-sensuous basic evidence, if there is any. For instance, we commonly judge of a man's disposition, or passing states of mind, by rapid, more or less automatic inductions based upon the formation of his features, the transient expressions of his face, etc. And yet we are inclined to claim that our conclusions are due to some kind of inner 'sympathetic' intuition. Again, two people in conversation will often both see the same thing without consciously noticing it, will both reason from it to some conclusion which they will both reach and utter at the same time; and they will then believe that this is a case of 'telepathy'. Again, moral intuitions are commonly supposed to be cases of basic non-sensuous immediate knowledge. But if our analysis has been correct, they are really founded on sense-experiences. For they ultimately result from observation of the behaviour of men's bodies. The wise men who were responsible for the generalization that the things of the mind are more valuable than the things of the body drew this induction

from observing how men act and speak, *i.e.* from observation of their fellow-men's bodily behaviour.

It is not meant to deny that moral, æsthetic, and religious intuitions have a high degree of validity. But they must not be set up as independent authorities claiming a superior right to be heard. For the reasons given they must not be allowed to contradict the plain evidence of the eyes and ears.

(5) The only way of testing the validity of unreasoned beliefs is to discover the evidence or unexplicit reasonings on which they are based, and then to examine this evidence or reasoning critically, applying to it the canons of good reasoning; in other words to rationally reconstruct them. It follows that there is no way of knowing *beforehand* what reliance ought to be placed on this or that unreasoned belief. Some way of doing this is of course what we should like to have. Since philosophers do, and probably always will, introduce their unreasoned beliefs into the foundations of their philosophies, some device for appraising their respective degrees of validity before they do this would be a godsend. There is, however, no way except the long way. Therefore it is the duty of philosophers, if they care for truth more than for beating down their opponents, to attempt the rational reconstruction of their intuitions before they found their philosophies upon them. This, it may be objected, is just what they now do when they try to support their convictions by reasonings which may have the appearance of being afterthoughts, when they try to find reasons for "what they believe on instinct". There is some truth in this. But if they realized that their intuitive beliefs are not immediate revelations, are the results of unexplicit reasoning processes which are extremely liable to error, they might at least be more careful and more honest. They would not, as they do now, *start out* by being absolutely convinced that their intuitions are true, and think that all they have to do is to support them by reasonings, specious or otherwise, *at any cost*.

(6) Only very general remarks can be made about the degrees of validity of the specific kinds of unreasoned beliefs roughly classified in section 3.

To begin with the universal factual beliefs which Russell calls instinctive. The belief that material objects exist when unperceived seems to me, for the reasons given earlier, to have no validity at all, unless we are prepared to accept its pragmatic value as validity.

The analogical argument for the belief in the existence of other minds is recognized by philosophers as not being conclusive.

But the degree of validity of an unreasoned belief will be the same as the degree which attaches to the same belief if reached by the same reasoning explicitly set out. Therefore our unreasoned belief in the existence of other minds cannot be regarded as certain. But this after all may be said of all conclusions based upon analogical or inductive reasoning. And I cannot see that our belief in other minds, whether reasoned or unreasoned, is any less reliable than beliefs in other well-founded inductions and analogies. I cannot see that it is any less reliable than our belief that the sun will rise to-morrow. In other words, it has an extremely high degree of validity.

Our second class of unreasoned beliefs, those which though not universal can claim the assent of most civilized peoples, were all moral intuitions. Although these cannot be placed, as some philosophers would like to place them, on a higher level than beliefs founded on the evidence of the senses, it seems to me that their degree of validity may in certain cases be very high. Particular moral precepts about particular kinds of actions will be less certain than the broader deeper intuitions such as that concerning the superiority of mental to bodily things. For the former are more dependent on the special and variable circumstances of different kinds of social groups. But if we take as our example the belief in the superiority of the things of the spirit, and if we analyse its meaning in the way that has here been suggested, we get, I think, the following result. The fact that the inductions on which it was based were unexplicit, unsifted, rough and ready, not safe-guarded by any scientific processes of verification, tells against a high degree of validity. But the facts that it is probably the result of thousands of years of human experience, of the unexplicit reasonings of many different minds in many different ages and countries of the world; that these minds working independently of one another yet tended everywhere to come to the same conclusion; that many of them—those who have come to be called the great moral teachers of the world—were in their time, and are still, recognized as being the most expert in moral matters, *i.e.* in making inductions about the probable consequences of different ways of living; that men still can and still do make similar observations on human beings and reach similar conclusions by more or less explicit inductions; that no one seems to have any intuition to the contrary; that (so far as I know) no competent observer of life has ever even attempted, by explicit scientific inductions, to prove the contrary—these facts seem to me to lend very great weight to this particular unreasoned belief. If anyone were to deny the truth of

it, I do not think, with Moore, that it would be "self-evident" that he is wrong. But I think that the best evidence we can get tends to show that he would in fact be wrong. Similar considerations would, I think, apply to many other moral intuitions.

Almost nothing can be said in general about the third class of unreasoned beliefs, those which are more or less personal to individuals, the particular metaphysical intuitions of a Bradley or a Whitehead, the various flairs, intuitions, and hunches of individual men, except perhaps that they are usually much less reliable than their owners suppose. Their degree of validity obviously varies enormously. All that we can ever do is to consider each case on its merits, testing it by the canons of rigorous reasoning.

III.—DISCUSSIONS.

HALL AND BERGMANN ON SEMANTICS.

PROFESSORS Everett W. Hall [1] and Gustav Bergmann [2] of the State University of Iowa have recently published papers discussing semantics as developed in a book of mine.[3] The following comments are intended to clarify some points which seem to me to have been misconstrued in these papers. These points concern the nature of pure semantics and, in particular, the problem of extra-linguistic reference.

1.

By 'semantics' I understand (following Alfred Tarski and Charles Morris) the theory of the relations between expressions in a language and their designata, *i.e.*, the objects, properties, relations, etc., to which the expressions refer. Among the semantical rules for a constructed symbolic language system (dealt with in what I call "pure semantics") there may, *e.g.*, be a rule (1) stating that a certain sign, say the letter 'a', designates Chicago, and another rule (2) stating that a certain other sign, say the letter 'P', designates the property of being large. Other semantical rules will tell us how to interpret a sentence if we know the designata of the signs occurring in it. Thus, *e.g.*, we shall learn that the sentence 'P(a)' of our symbolic language system says that the object designated by the sign 'a' has the property designated by 'P'; hence, in view of the rules (1) and (2), 'P(a)' says that Chicago is large. It can then easily be shown that the concept of truth with respect to the symbolic language in question can be defined in (pure) semantics. In order to avoid some technical complexities, I will not give here the definition in its general form. It seems plausible that, if the definition is constructed in a suitable way, it will lead to the following result as a particular instance; the sentence 'P(a)' is true if and only if Chicago is large, that is to say, if the physical world we see around us is such that Chicago, that well-known complex physical thing consisting of houses, streets, etc., is large. So much is stated in (pure) semantics. However, the task of finding out whether or not Chicago is large, and thereby whether or not the sentence 'P(a)' is true, is not a task of semantics but of empirical science. Semantics has the task not of fact-finding but of interpreting language. Although limited to this task, semantics does and must refer to extra-linguistic entities, *e.g.*, physical objects, their properties and relations. This is clear from the examples mentioned; rule (1) refers on the one hand

[1] E. W. Hall, "The Extra-Linguistic Reference of Language, Part II: Designation of the Object-Language", MIND, vol. liii, 1944, No. 209, pp. 25-47.

[2] G. Bergmann, "Pure Semantics, Sentences, and Propositions", *ibid.*, No. 211, pp. 238-257.

[3] R. Carnap, *Introduction to Semantics*, 1942.

to a certain sign, but on the other to the thing Chicago, and it states a certain relation between the sign and the thing; analogously, rule (2) refers not only to a sign but also to a certain physical property, *viz.*, that of being large; and the semantical statement given above which states a truth-condition for the sentence 'P(a)' refers on the one hand to this sentence, hence to something within the language, but on the other hand to the state of affairs (or possible fact or what else you want to call it) of Chicago's being large, hence to something extra-linguistic.

In my book I have expounded in detail the nature of semantics as briefly outlined above. I have especially emphasized its distinction from syntax (Morris' "syntactics"); while the latter discipline deals only with relations among expressions in a language and thus with an entirely intra-linguistic subject-matter, it belongs to the essential characteristics of semantics that it refers not only to language but also to extra-linguistic matter. The book does not only state and emphasize this characteristic in abstract terms, but also illustrates it by a great number of examples of semantical rules for various language systems, such as the two rules given above, and of semantical statements concerning truth, as in the example above, and other semantical concepts.

2.

In view of these facts it will seem understandable that I felt some surprise—or should I rather say "amazement"?—when I saw that Professor Hall in the paper mentioned comes to the conclusion that (pure) semantics, as I construct it and apply it, does not and cannot refer to anything extra-linguistic, but deals only with relations between expressions in a language, and that all appearance to the contrary is merely due to some careless, misleading formulations of mine. This result is not stated as an incidental remark, but rather as the chief thesis in a lengthy discussion (*loc. cit.*, pp. 25-36) of my book, reiterated again and again in various forms. I will give only a few examples of formulations in which Hall ascribes to me views which are just the opposite of my conception of semantics. (The italics and insertions in parentheses are Hall's; the insertions in square brackets are mine.) "Carnap means to rule out all consideration of the designation of zero-level expressions [*i.e.*, expressions in the object-language, like 'a' and 'P'], *i.e.*, of any expressions save the names of expressions" (p. 25 f.). "Pure semantics cannot say anything about the designation or truth of any expression at zero-level [like 'a', 'P', 'P(a)'], for to do so would involve an assertion (however general and abstract) about matter of fact as well as about language, and would therefore be descriptive. This conclusion (though explicitly mine) seems to be fairly drawn; that is, Carnap seems to accept it" (p. 27). "Designation is wholly a relation *within* [the language system] *S*, therefore within a language (semantical) system, therefore not a relation of language or linguistic

elements to extra-linguistic fact" (p. 27). "The key idea is that designation is to obtain between two expressions in a metalanguage. . . . No extra-linguistic reference of language is considered" (p. 27, footnote 1). "Carnap . . . means that pure semantics, in studying truth, is confined to propositions asserting the (truth-)equivalence of sentences at different levels" (p. 28). "He is saying not that we can predicate truth only when we know the relation of a zero-level sentence [like 'P(a)'] to fact, but quite the opposite, that we can properly predicate truth on the basis of linguistic (semantical) rules alone" (p. 29); compare with this the following passage in my book (p. 33): "Thus we have found a truth-condition for [a certain symbolic sentence] in [the language system] S_3 as determined by the [semantical] rules of S_3. But these rules do not suffice to determine the truth-value [*i.e.*, truth or falsity] of [that sentence]. In order to find this, we must know certain facts in addition to the rules. This would lead us outside of semantics into empirical science, in this case into geography." Hall says further (p. 34): "I think he [Carnap] would simply admit that semantics cannot escape the linguacentric predicament. Semantics are arbitrarily or by conventional division of labour confined to language (in its designative aspects) as abstracted from all matter of fact, he might claim."

Once Hall is convinced that I intend to restrict semantics to intra-linguistic matter and, in particular, the relation of designation to pairs of expressions, he proceeds to give a new and strange interpretation to those of my formulations which manifestly speak about extra-linguistic entities; he believes himself compelled to reject the natural, literal interpretation because this would violate my alleged principle of restriction. He says (p. 30): "Thus the designata of '*u*' and '*p*' might be thought to be extra-linguistic. . . . This, however, would clearly presuppose that expressions . . . [can designate] the same extra-linguistic entity, and is thus subject to the criticism of that assumption we noted above, namely that it violates Carnap's restriction of designation to pairs of terms." However, there are places where my references to objects, properties, and other extra-linguistic entities are so clear and explicit that not even artificial reinterpretation will do. In these extreme cases there is no other way left than suggesting a "correction" of my formulations. For instance, Hall quotes (p. 29, footnote 3) the following sentence from my book (p. 24): "The object designated by . . . has the property designated by - - -" (where technical names of signs stand at the place of the dots and dashes). One might think it difficult to maintain in view of this sentence that my semantics never refers to objects and properties. But not so for Hall; he simply assumes that I didn't mean it. He says that the sentence quoted "is a little misleading. In place of 'object' 'object-sign' would be less misleading, and in place of 'property' 'property-sign'." Consequently, he translates the sentence quoted into "the object-sign has the predicate-sign attached to it"; and this, he thinks, is what I

actually meant. The same method is again applied later (p. 32). Hall quotes my heading (p. 128) "Procedures of the second kind: the elements of L-ranges are extra-linguistic". How is this compatible with the thesis that semantics never refers to anything extra-linguistic? Simple again; you have merely to assume that I meant by the term 'extra-linguistic' something intra-linguistic. Hall says: "Carnap's heading . . . is misleading. Instead of 'are extra-linguistic' it should read 'are symbols whose designata are extra-linguistic' or 'are symbols of the zero-level'." Thus he has again proved his point that my semantical theory deals exclusively with expressions and never with objects or properties or anything outside of language. His argumentation is quite irrefutable. I recommend to all philologists this far-reaching method for ascertaining what an author actually means; if we were to apply it to his own statement (p. 39) "With Russell's solution of his problem I disagree wholeheartedly", we could readily establish that he actually meant to say that he agreed completely.

How is this profusion of thorough misunderstandings to be explained? I wish to emphasize first of all that they are certainly not due to careless reading or rash and dogmatic conclusions on Mr. Hall's side. On the contrary, the discussion in his paper is marked by careful and conscientious procedure. He says repeatedly that the interpretation at which he arrives is his impression, his feeling, his conclusion. He admits that he may be mistaken in this interpretation. Several times he considers alternative interpretations (the chief one of these is outlined on p. 26; I agree with him that this is not what I meant); and only after lengthy deliberation of reasons pro and con does he accept his interpretation.

As I see it, the explanation is to be found in the fact that Mr. Hall makes one initial mistake, slight in appearance but serious in its consequences. After this, all his derivations, careful though they are, lead him to more and more errors. Even the one original sin cannot be ascribed to carelessness; it consists in continuing the tradition of a certain confusion of which some of the best men in the field of logic have been guilty.

One of the fundamental distinctions which must be made when speaking about language is that between *using* an expression and *referring* to it. When I say: "Chicago is a large city", I am using the word 'Chicago'; I am referring not to this word but to the thing whose name is this word, *viz.*, the city of Chicago. On the other hand, if I wish to refer to the word 'Chicago', *e.g.*, in order to say that it consists of seven letters, I must use not the word 'Chicago' but a name for it. I may, *e.g.*, introduce 'c' as a name for the word 'Chicago'; then I can express the above assertion by means of the sentence: "c consists of seven letters". A simpler and more customary procedure for constructing a name for a linguistic expression consists in simply including the expression itself in quotation marks. Thus we may say: "'Chicago' consists of seven letters".

It does not matter which method for constructing a name for an expression we use; the essential point is that we use a name for the expression and not the expression itself whenever we intend to speak about the expression. If we do otherwise (*e.g.*, if we say: "Chicago consists of seven letters"), then the expression is used ambiguously, and many confusions are apt to arise. At the first blush, the distinction seems a minor matter and to urge it seems rather pedantic. However, it is in fact a very serious matter. Frege[1] first emphasized the necessity of the distinction; and recently Tarski,[2] Quine,[3] and I[4] have repeatedly called attention to the same point. It has been shown that the neglect of the distinction in the Introduction to *Principia Mathematica* has made many points obscure and ambiguous and in some cases has even caused serious confusion and errors. Certain later authors, who followed the example of Whitehead and Russell rather than that of Frege, have created still more confusions and errors.

The one fundamental mistake made by Hall is just the neglect of this distinction between using an expression and referring to it. Consider the semantical rule: "'a' designates Chicago". The rule says that the relation of designation holds between two entities, which are specified in the rule. What are these two entities? We find in the rule at the first argument-place (*i.e.*, at the place preceding the relation-word 'designates') the name of a sign. (In the above formulation this name is built with the help of quotation marks; in my book, I use instead certain German letters which are introduced as technical names of signs; this difference is inessential, as mentioned before). Therefore, what is referred to in the first place is the sign whose name occurs, *viz.*, the first letter of the Roman lower case alphabet. At the second argument place (*i.e.*, following the word 'designates') we find the word 'Chicago'. Therefore, what is referred to in the second place—and this is the decisive point—is the entity whose name occurs at the second place; this entity is not the word 'Chicago' but the physical thing Chicago. Thus we see that the rule as formulated above can be interpreted in only one way; it says that (in the symbolic language system in question) the first letter of the alphabet designates (that is, is used as name of) the thing Chicago. In addition to examples of rules of designation like the one above, I have given in my book a general description of the form of these rules (D12-B, p. 53); this description says that, as we have seen in the example above, the name of an expression stands at the first argument-place (*i.e.*, the place preceding the relation word

[1] G. Frege, *Grundgesetze*, vol. i, p. 4.

[2] A. Tarski, *Introduction to Logic*, § 18; see also the useful exercises *4 and *5 on p. 65.

[3] W. V. Quine, *Mathematical Logic*, § 4; see also the footnote on p. 283. —Quine, "Whitehead and the Rise of Modern Logie" (in *The Philosophy of A. N. Whitehead*, ed. A. Schilpp, pp. 127-163), see pp. 142, 145, 151 f.

[4] Carnap, *Logical Syntax*, § 42. *Introduction to Semantics*, pp. 236 f.

' designates '), and another expression (above the word ' Chicago ') stands at the second argument place ; and it says furthermore that the second expression is the translation into the metalanguage (here, English) of that expression whose name stands at the first place. (In our example, the word ' Chicago ' is the English translation of the sign ' a '.) To sum it up, the rule *uses* at the two argument-places a name of a sign and an expression ; it *refers to* a sign and a thing. By confusing use of an expression with reference to an expression, Hall arrives at the erroneous belief that designation is a relation between two expressions. (Rules of designation of the form " ' Chicago ' designates Chicago " or " ' a ' designates a " seem to Mr. Hall still more puzzling. Rules of this form, it is true, do not convey much information ; but the decisive point is that they too contain extra-linguistic reference. The first rule says that the word ' Chicago ' designates the thing Chicago, and the second says that the sign ' a ' designates the thing a.) Hall's inference is just as fallacious as the following would be : " The sentence ' Jack is older than Peter ' contains two names of persons ; therefore this sentence speaks about names and not about persons". It seems to me that all of Hall's misinterpretations stem from the one basic mistake I have tried to indicate. This mistake clearly underlies the whole first part of his discussion (pp. 25-28). It is most explicitly formulated at the following two places. 1. Discussing my general description of rules of designation (D12-B), Hall says (p. 27, footnote 1) : " The key idea is that designation is to obtain between two expressions in a meta-language when the first is the name of an expression of which the second is a translation. . . . But what is named and translated is a linguistic expression. No extra-linguistic reference is considered." 2. In my book I have sometimes used, instead of the formulation " ' P ' designates the property of being large " (*e.g.*, p. 24), the shorter formulation " ' P ' designates large " (§ 12) ; the latter formulation is more convenient, though admittedly not in accordance with customary English grammar. This shorter formulation does not involve any ambiguity because the use or non-use of quotation marks makes always clear whether we speak of the word ' large ' or of the property large (*i.e.*, the property of being large). Hall makes here the same mistake as before ; he ignores the fact that, although ' large ' is certainly a word, large cannot possibly be a word. He says (p. 28) : " Carnap's example is : ' Gross ' in German designates the same entity as ' P ' in his [language system] S_3, *viz.*, large. But here large must be a word (in English) not a property of fact. For, ' designates ' in " ' Gross ' designates large " and " ' P ' designates large " is defined by pairs of expressions (' Gross ', large ; ' P ', large) the first members of which are names of expressions of which the second are translations, hence the designatum in each (large) is an expression. . . . So again we find that only names of expressions are allowed to designate, and they can only designate expressions." Thus we see in both passages here quoted, how the

mistake which I have called the basic one leads to further errors, first to the conclusion that the designatum in the case in question is an expression while in fact it is an extra-linguistic entity, and then to the general conclusion that semantics does not make any reference at all to extra-linguistic matter. It could easily be shown that all the numerous other misunderstandings of my formulations and the independent errors in Hall's paper, of which a few have been quoted earlier, grow out of the same root.

3.

In the article mentioned above, Professor Bergmann discusses the nature and function of pure semantics as developed in my book. Some points in his discussion are interesting, other parts I find hard to understand. Unfortunately, his discussion also involves some serious misunderstandings and errors. Since they are similar to those of Hall discussed above, I can be brief here. Bergmann gives the following characterization of pure semantics, which seems to me, to say the least, very misleading: "Meta-languages of the kind . . . are called *semantical metalanguages*. Pure semantics is the study of such metalanguages formally considered" (p. 246). I should prefer to say that pure semantics is the study of object languages, not of metalanguages, and further, that it is not a formal study but a study of interpretation. I am not certain whether the passage quoted is merely an inadequate formulation or whether it shows an actual misunderstanding of the nature of pure semantics. The following points seem to make the latter alternative more likely.

Bergmann says: "The one domain of these predicates [the relational predicates of designation] consists of the names of the expressions of the object language, the other domain, of the metalinguistic names of the referents [designata] of those expressions" (p. 245). This seems to me a confusion between names and the entities named; in fact—as earlier explained in a discussion of the rule "'a' designates Chicago"—the first domain contains expressions of the object language (*e.g.*, the sign 'a'), not their names, and the second domain contains the designata of those expressions (*e.g.*, the thing Chicago), not their names. On the basis of this misunderstanding, Bergmann arrives, in a way similar to Hall's, at the absurd conclusion "that pure semantics does not deal with the *extralinguistic* referents, the designata" (p. 248). A confusion of the same kind as that mentioned above is to be found in the following passage (p. 252): "In the statement:

(3) 'Fido is a dog' is true if and only if (Fido is a dog). Carnap refers to the quoted expression as a sentence, while the bracketed expression is called a proposition." In fact, presumably like most other logicians, I call the quoted expression a name of a sentence and the bracketed expression a sentence; in the book under discussion, I do not apply the term 'proposition' to sentences or to any

other expressions. Bergmann believes erroneously that my terms ' proposition ', ' individual ', and ' designata ', " if they are to occur in pure semantics, do not refer to any extralinguistic referents " (p. 252). The fact that my choice of these terms and the formulations in which I use them show clearly that I intend to refer to extralinguistic entities is explained away by Bergmann, just as by Hall, by the assumption that my formulations contain " a misleading terminology and certain ambiguities " (p. 238, similarly p. 252). This assumption, however, is merely the outcome of Bergmann's misunderstanding, as the example of the term ' proposition ' shows.

4.

My remarks on Hall's and Bergmann's papers may be summed up as follows.

(*a*) The two authors fail at some crucial points to make the necessary distinction between use and mention of an expression, in other words, between an expression and its name.

(*b*) Consequently, they interpret what I say about rules of designation as if the expressions used in these rules were expressions mentioned ; thus they come to the erroneous conclusion that the designatum is always an expression.

(*c*) Consequently, they believe erroneously that pure semantics does not refer to extra-linguistic matter.

When I say that certain of my formulations of semantical statements and rules were intended to refer to extra-linguistic entities, I expect that Hall and Bergmann will concede this point, because it is generally agreed that an author is the final judge in a question as to what he meant to say. On the other hand, an author is not the final judge in a question as to what he did actually say, *i.e.*, whether he succeeded in formulating what he intended to say. (For instance, when Hall writes (p. 26, line 17) : " that A designates that matter of fact ", I wonder whether he did not actually mean to say : " that ' A ' designates . . . ".) Therefore, Hall and Bergmann have to decide whether they will agree with me when I say that the interpretation of my formulations as referring to extra-linguistic entities is the only correct one and that their interpretations are based on a confusion of use and mention of an expression. I have often found that for those who have acquired the habit of this confusion it is very hard even to recognize it. But I think that all those who are free of this confusion, in other words, those who distinguish clearly between an expression and its name, will agree with me on this point.[1]

[1] Just after having written my comments, I find that Ernest Nagel comes to a similar conclusion. In his review of Hall's paper (*Journal of Symbolic Logic*, vol. ix, 1944, p. 46), he says : " This is an unfortunate misunderstanding of Carnap's language, and in consequence the entire discussion of semantics suffers from a fundamental *ignoratio elenchi*."

RUDOLF CARNAP.

HEMPEL'S PARADOXES OF CONFIRMATION.

IN MIND for January, 1945, Mr. C. G. Hempel introduces what he calls the "paradoxes of confirmation". They are as follows. We make the plausible assumption that "an object satisfying two conditions C1 and C2 should be considered as confirming a general hypothesis to the effect that any object which satisfies C1 also satisfies C2" (Hempel, p. 14). If, then, we take the two general statements

S1. All ravens are black,

and S2. Whatever is not black is not a raven,

then S1 is confirmed by the fact that a given raven is black, and S2 is confirmed by the fact that a given object which is not black (say, a red herring) is not a raven. But S1 and S2 are logically equivalent, and whatever confirms S2 also confirms S1. Therefore the red herring, or any other object which is neither black nor a raven, confirms the hypothesis that all ravens are black. This is the first paradox. The second is similar. For

S3. Anything which either is or is not a raven is either no raven or black

is equivalent to S1 and S2. Therefore any object which is not a raven, whether it is black or not, since it confirms S3, also confirms S1. These consequences are paradoxical, but, Hempel argues, they must be accepted as valid.

It seems to me plain that they are not only paradoxical, but false. No case, and no number of cases, of red herrings, white elephants, etc., can in any degree lessen the likelihood that there are some ravens which are green; and no person in his senses would present his observations on herrings and elephants as evidence for the assertion that all ravens are black. To avoid the paradox, there are two possible courses. One is to deny that, for the purposes of the argument, S1 and S2 are equivalent; the other is to deny that the red herring is evidence either for S1 or for S2. I take the latter course. It involves the rejection of Hempel's initial assumption that "an object satisfying two conditions C1 and C2 should be considered as confirming a general hypothesis to the effect that any object which satisfies C1 also satisfies C2"; for the red herring satisfies the condition of not being black and also the condition of not being a raven, and yet, in my view, does not confirm the hypothesis that whatever is not black is not a raven.

It is important to be quite clear what is meant here by "confirm". Three relations between a particular fact and a general hypothesis are in point.

(1) The fact may be an "instance" of the hypothesis. Any black raven is an instance of S1, and any non-black non-raven is an instance of S2, and everything in the world except a non-black raven

is an instance of S3. So in this respect the three statements are not equivalent.

(2) The fact may be consistent or inconsistent with the hypothesis. Anything except a non-black raven is consistent with S1, S2 and S3.

(3) The fact may be evidence for or against the hypothesis, *i.e.*, may make it more or less probable. This is the sense in which Hempel is using "confirm". And the matter in dispute is whether a fact which is an instance of a hypothesis, or again a fact which is consistent with a hypothesis, is thereby evidence for that hypothesis. Now to pose the question in this way tempts us to suppose that (3) is a relation of the same sort as (1) and (2). But it is in an important respect different. For whereas (1) and (2) are relations which hold simply and directly between the general hypothesis and the particular fact, and in order to know whether a particular fact is an instance of, or is consistent with, a general hypothesis all that I need to know is the fact and the hypothesis, (3) is a much more complex relation: no single fact, taken by itself, is evidence for any generalisation at all, but all inductive argument depends on the comparison of different cases. Taken by itself, without reference to what we know of other ravens, and without any presuppositions regarding the constancy of colour in birds, one black raven gives us no ground for believing that all ravens are black.

Yet the existence of black ravens is certainly evidence, and indeed the main part of the evidence, for S1; *i.e.*, S1 is established by observing instances of it. Why is this not true of S2? The only difference between S1 and S2 is that in S1 the conditions specified are positive, and in S2 they are negative. This difference is relevant; a statement that whatever has the character C1 also has the character C2 is not established in the same manner as a statement that whatever has not C1 has not C2. The reason is as follows. If S1 is false, *i.e.*, if there are ravens which are not black, it is *a priori* unlikely that a collection of ravens taken at random will consist entirely of black ones. Hence a series of ravens which are all black gives us a situation which is unlikely to occur if S1 is false, and therefore provides evidence in favour of S1; the strength of the evidence varies with the extent and the randomness of the collection; and again, we require a good deal more evidence to favour S1 as against the hypothesis "Most but not all ravens are black" than as against the hypothesis "Most ravens are green". It is clear that the occurrence of one black raven is not improbable even if S1 is false, and therefore provides no evidence for it. Now what about S2? Whether S2 is true or false, it is not in the least improbable that a collection of red herrings or white elephants should contain no ravens; and our failure to find any ravens in such a collection is not, like our failure to find red ravens among a collection of ravens, an improbable occurrence on the assumption that S2 is false, and is thus no evidence in favour of S2. Thus the herring, whether by itself or as one of a collection of herrings, is irrelevant.

All the same, do we not sometimes use something which has neither C1 nor C2 as evidence that whatever has C1 has C2? Certainly. But it is not everything which has neither C1 nor C2 which is capable of providing such evidence. The evidence which can be so used is that drawn from those kinds of things among which we might expect to find things which have C1 but not C2 if there are such things. Thus (to take Hempel's other example) if among a number of substances which have similar properties some both contain sodium salts and burn yellow and others neither contain sodium salts nor burn yellow, the absence of both characters in the negative instances reinforces their presence in the positive instances; but to examine a piece of ice, which is not at all like a sodium salt, gives you no addition to your evidence (as Hempel seems to admit). And this evidence is valid only in conjunction with the positive instances; if we know merely that none of our substances either burns yellow or contains sodium salts we have no evidence about the behaviour of sodium salts.

In sum: the evidence for an inductive generalisation always consists of several facts taken together, and to suppose that each of these facts taken separately must possess some weight as evidence is a mistake; the relation of being-evidence-for a general hypothesis is of a different kind from being-an-instance-of or -being-consistent-with such a hypothesis, and is not to be dealt with on the same lines; and a fact may be consistent with a hypothesis, or even be an instance of it (especially where the instances have no common character, as is the case with S2 and S3), without being evidence for it.

C. H. WHITELEY.

A NOTE ON THE "ARGUMENT FROM ILLUSION."

I AM much impressed by the thorough and able analysis which Mr. C. Lewy undertook in a refutation of Mr. Ayer's "argument from illusion" ("On the Relation of some Empirical Propositions to their Evidence", MIND, October, 1944), yet I am not entirely convinced for two reasons concerned with Mr. Lewy's treatment of his proposition (4), "I have all the evidence which I do have that there is a desk in my room, but there is no desk in my room", and I shall state them as briefly as possible in hope that Mr. Lewy has a ready answer and in order to elicit it from him.

In his treatment of (4) Mr. Lewy asserts that it is a fact that one can imagine the following situation:

"Now suppose that having all the evidence which I do have that there is a desk in my room, . . . I look away for a moment, and then, looking back again, can see the desk no longer. I look again, and again see nothing; I try to touch the place where I thought the desk was standing, but can touch nothing. I bring in other people, all my previous witnesses, and they, too, no longer get any 'deskish' sensations" (p. 294).

Mr. Lewy then adds: "*This* we can imagine: we have just done so." But I think there is an ambiguity in this conclusion. For to imagine a desk we may either simply entertain an image of a desk or entertain such an image while conceptually referring it to an actual material thing, for example, to the desk in my study. If we imagine the desk in the latter way, the imaginary setting must be consistent with whatever is implied by the concept of a material thing. Thus we can imagine an *image* of the desk to suddenly disappear as described in the quoted passage. A dream-image of the desk may vanish in a dream, and an image of my desk vanishes in my imagination as I follow Mr. Lewy's description in applying it to myself. But I do not think that I can imagine the desk itself, as contrasted with an image of it, to disappear in the way in which I have imagined its images (free from conceptual reference) to disappear. I suppose I cannot do this because the meaning of the word "desk", as opposed to "an image of a desk", entails the insignificance of such combinations of words as "the desk (as a physical object) suddenly ceased to exist without any physical cause". Yet, I am afraid, Mr. Lewy intended his "imaginary situation" to involve not only images but conceptual reference to actual things. If so, it would seem that his description of the imaginary situation does not make sense. And if we conclude that his description does not make sense, our conclusion would seem to be favourable to Mr. Ayer's argument from illusion, for Mr. Ayer may say that the argument is, in effect, equivalent to the observation that one can never tell whether the concept "desk", as opposed to "an image of a desk", happens to be exemplified in any particular situation; or, to put it differently, that we can never be sure whether we perceive more than an image of a desk, even if, in fact, we actually perceive the

desk itself. For consider two statements. (1) There was a desk here a moment ago, but it suddenly and miraculously vanished. (2) The fact that there was here an image of a desk a moment ago but that it suddenly vanished means that the image was an illusion, or a dream, or a hallucination. It would seem that, if miracles were admitted to endow (1) with sense, there could be no evidence for or against (1) which would not equally be evidence for or against (2), so that the meanings of the two statements would be, for all practical purposes, indistinguishable.

To give the second reason for my doubts, I must assume, for the sake of argument, that what I said in the preceding paragraph was wrong, and that Mr. Lewy's imaginary situation must be accepted without qualification. Now Mr. Lewy asserts that only verbal reaction to such a situation, should it ever happen to occur, can determine whether the situation is relevant to the question of the self-consistency of (4). And on page 297 he admits the *possibility* of some such verbal reaction as the statement "So there wasn't a desk here after all—we were all victims of a long and very consistent hallucination". But the possibility of making this statement implies the possibility that Mr. Lewy's imaginary situation is relevant to (4). And if the imaginary situation may be relevant to (4), the evidence for the existence of the desk, as stated in (4), does not entail the existence of the desk; and therefore (4) is self-consistent, contrary to Mr. Lewy's conclusion to the effect that one can say neither that (4) is self-contradictory nor that it is not. My objection, as thus stated, occurred to me while I was reading Mr. Lewy's page 297, and I was surprised to find him raise on a subsequent page (p. 311) an objection, to his own conclusion, which seems to be allied to, if not essentially identical with, my point. His definite statement on this matter would be of particular interest to a logician; for if his previous conclusion, that (4) is neither consistent nor inconsistent, stands unchallenged, it is not only an epistemological finding, but a proof that the law of the excluded middle is not a valid theorem: the proposition "(4) is self-contradictory" is neither true nor false.

A. USHENKO.

IV.—CRITICAL NOTICES.

Man and his Works. By E. L. THORNDIKE. (William James Lectures, 1942-43). Cambridge, Mass., Harvard University Press; London, H. Milford, 1944. Pp. 212. $2.50, 14s.

PROFESSOR THORNDIKE plunges at once into a discussion of Nature's gifts to Man. He prefers genetic language, and calls the gifts in which he is most interested 'genes'. Others have called them 'tendencies', 'predispositions', 'instincts', and even, sometimes, 'faculties'. Whatever name they are given they are always supposed to have some special concern with action. They are elements, or groups of elements, in the internal constitution of animals, including man, which require stimulation from outside to make them do anything, but which, so far as they themselves go, owe nothing else to the outside world. People who call them 'tendencies', or something like that, generally like to think of them as limited in number, and they often make short lists which somebody else at once begins to dispute. Thorndike, calling them 'genes', likes to think of them in their thousands, and this certainly avoids one rather fruitless squabble. He thinks of thousands of genes co-operating in thousands of ways. When I got to this point I wondered whether he also thought of them as fighting one another sometimes. For it appears to me that they do, and that this may be very important indeed. But the general drift of the whole book shows that he is all for co-operation and all against mutual interference, and that is certainly an admirable characteristic in him, though it may be held to limit his theories.

The genes are not only very numerous, but they do not operate as entities or fixed groups, and so allow for an immense variability in original behaviour; they are not modifiable from generation to generation; they require outside stimulation through the sense organs, and the ones that contribute most to the works of man are those which lie at the basis of speech and tool-using.

In all this there neither is nor pretends to be anything original. But there are some exceedingly ancient difficulties. The great bulk of man's behaviour now consists of learned activities. How, if there ever is a stage in which only genes are operative, does such learning begin? Thorndike does indeed contemplate certain types of gene-determined behaviour as 'insufficient'. It is not very clear for what they can be insufficient unless some additional assumptions are made about genes besides those that are mentioned. But in any case he uses the consideration, correctly, only to distinguish between behaviour which seems to express some 'general tendency' and behaviour which requires a specialised mechanism, both being equally and entirely gene-determined. He is "confident" that the genes "provide no direct and primary responses to any mental images, or ideas of concrete things, or judgments about them, much less to

general or abstract ideas and judgments about them." Difficulties accumulate when he begins to talk about speech and tool-using: "Man by nature moves the tongue and other mouth parts to produce a great variety of sounds, and has pleasure in doing so. These movements, in connection with certain intellectual powers, can produce meaningful speech." The "certain intellectual powers" are a bit of a mystery. They come in with a bang, casting no shadow before them and bearing no credentials with them. They cannot be special genes because their objectives are just the sorts of things that no gene can do anything about. If they are to be accounted for at all, the conditions of their appearance must belong, so far as Thorndike is concerned, to a study of learning process.

To this study he devotes the second chapter, which is vital to very nearly everything else in the book. A man's mind is depicted as an extraordinarily elaborate network of connections. To describe anybody fully we have to make a list (Thorndike is fonder of lists than anybody else I have ever known. Pretty well all his experiments present people with lists. If he has any problem to solve he first makes a list. Most of his illustrations are enumerations. Therefore, perhaps, one should not be surprised to find that man himself is only a list) of all the situations "he might encounter and be sensitive to and lists of all the responses he might possibly make to each, and statements of the probability that any of the situations would evoke any of the responses in that person". If S is situation and R response, a man is "the total of all his S-R probabilities", where "total" obviously means "list". It comes with a shock to find Thorndike roundly asserting that this is true, and all of the truth, whatever conceivable kind of theoretical foundation psychology may have. All that environment does is to change the values of the R probabilities up or down, and all the wrangles about different sorts of psychologies are quarrels about the number and character of the "forces" that do this.

Thorndike dubs his view "connectionist" and says its distinguishing feature is that it uses only two such "forces"—*repetition* and *reward*.

Fortunately Thorndike does not make much play with repetition. He says that strictly it ought to be called "occurrence", because the occurrence of any response very slightly increases the probability of its recurrence. In the setting of general theory provided, this does not come to much. For response, to Thorndike, is some sort of movement. The exercise of any sort of movement produces a degree of pleasure. Pleasure is a kind of reward, indeed the only basic kind there is. It seems as if "mere occurrence" can signify nothing at all, and as no carefully controlled experiments ever have demonstrated that sheer repetition is of any importance in learning, this need not occasion any worry. There is, then, one "force" only to account for learning, the "force" of reward.

Perhaps two of the many questions that can be asked at this point are the most interesting. How does reward work? and, Is

XUM

Thorndike successful in his attempt to show that no other influence operates to promote and direct learning?

Thorndike's answer to the former question is perfectly definite. Reward works by strengthening connections, chiefly between the response and the situation to which the response is made, but also to a smaller degree, between that response and neighbouring situations. Any psychologically trained reader will know that Thorndike's conclusion is based upon many years of patient, ingenious and original experiment. All the experiments have a strong family likeness, and it is worth considering briefly the one described in this chapter.

A list of 135 words was made and to each word one of the numbers 1 to 5 was assigned, there being, of course, no reason inherent in the word why it should be attended by one number rather than another. The list was then read aloud to many and different kinds of auditors who were asked to guess and record the correct number in the case of every word. At irregular intervals the correct number of a particular word was also stated. If the subject happened by chance to have hit on the right number, this statement constituted his reward. Invariably, confirmed numbers reappeared in subsequent readings with a frequency much greater than could be accounted for by chance and the same numbers were also assigned to near-by words in the list more often than their random use would explain.

The conclusion seems plain and just. A connection between giving a particular number in response to special words has been strengthened by confirmation, or reward. Thorndike immediately generalises on a large scale. The normal man experiences confirming reactions in their dozens every minute of every working day. They all work the same way.

"To each phrase that I speak you respond by a certain meaning. If this meaning satisfies you by making sense the connection is confirmed. Similarly, in reading a page a satisfying flow of sense is confirmed step by step."

At this point it seemed to me that a simple trial of these views might be made. It could not by itself be conclusive but it might suggest whether further tests would be worth while, and how they might be planned. I read on to the end of the chapter and closed the book. Thorndike's exposition is beautiful in form. With very few exceptions I think I fully appreciated the meaning of the sentences and paragraphs, so that they were satisfying in that sense. In a great part I also approved the meaning so that a further, and perhaps a different sort of, satisfaction had been present. It seemed to me that if the connectionist theory is correct I ought to be able, to a considerable extent, to repeat the language, particularly if I could begin with a few "key" words. But no, nothing of the sort happened. I could write down the argument; and other people who had read it, or were familiar with it, could easily identify it as derived from Thorndike. But the words, the sequence, the style were different.

The truth is, I think, that a connectionist theory of learning seems to be adequate in proportion as the items of a situation are more separated and individualised than is usual in daily life, and the responses attached to them more arbitrary. In most learning experiments this is just what does happen, and when, as to an altogether preponderant extent in Thorndike's experiments, they consist of parallel lists, it happens in its most extreme form. The conclusion is concealed in the design of the experiment and not in the natural facts.

It is more difficult to state briefly how reward, in so far as it operates at all, does work. But it seems as if it must be *either* by strengthening specific connections *or* by stimulating a mass of material whose items may vary freely in sequence, form and individual characteristics while yet their general significance remains constant. The second kind of learning appears to be very much more frequent under everyday conditions than the first. Perhaps it might be urged that the second type is properly described as the sort of instance in which the odds against any particular response are extremely small, and those in favour of a large number of differing responses are about even. Very few psychologists would, I think, agree with that.

Is reward the only "force" operating in learning? This is a hard question to answer and though Thorndike's exposition seems to amount to a strong assertion that it is, I am still uncertain whether he really does think so. Partly this is due to the extreme ambiguity of the word "reward". Reward can, in fact, be applied to almost any state of mind that follows action, within, possibly, certain rather wide temporal limits. The only reward of truly gene-determined action is pleasure. But Thorndike says little about pleasure. The words he prefers are "confirming reaction", "satisfyingness" and even "belongingness". The trouble about all these is that they stand for objectives which, by his hypothesis, no genes can possibly deal with. Presumably the mysterious "intellectual powers" must come in again. Even on the gene level, pleasure is not the only reward. Thorndike specifically dissents from any theory of hedonistic selection. This suggests the view, with which, so far as I can gather, he would agree, that different groups of genes have different degrees of potency. In that case the genes will not only co-operate, they will also interfere with one another. Obviously, however, such conflict can do nothing to effect a change of the principle of solution from pleasure to, say, belongingness, unless it brings some new determinant on the scene. It may be that we are to suppose that when two sets of genes of differing potency collide, the "intellectual powers" are awakened, or, in some sense, originated. But Thorndike does not say so, and if he did he would have at least two kinds of reward, on two such different levels that nothing could combine them into one.

Apart from this it seems to me that "confirming reaction", "satisfyingness" and "belongingness" are genuinely different

rewards, each appropriate to its own special circumstances and leading to different results.

What about punishment? Most people continue to think of this as having a genuine effect upon learning, different from that of "reward", negative in an immediate sense, though, according to some views, positive in a remote sense. In a connectionist theory, if this view is correct, punishment must first work either by preventing a connection from being made, or by weakening a connection that has already been made.

Thorndike, therefore, as most people will know, carried out a large number of multiple-choice experiments in human learning. The punishment he used was to tell his subjects when they made the "wrong" choices. "What was my surprise", he writes, "to find that a connection made and punished was more likely to be made again than if it had not been made at all!" He generalises that punishment has no direct effect in producing or tending to produce an avoiding reaction.

Once again, I believe, we have a result which is more a matter of the design of the experiment than of anything else. Thorndike's multiple choices were usually very numerous, far beyond what any normal person could take into his "span of apprehension". The mildness of the punishment may not be very important, for he analyses the results of a number of animal experiments and arrives at the same conclusion, when the punishment appears to have been anything but mild. But his method one way and another certainly reduced punishment to a sort of minimal significance. In some so far unpublished experiments carried out at the Cambridge Psychological Laboratory, Miss A. Dand repeated Thorndike's procedure, except that she reduced the number of choices to a range lying within or just within the "span of attention" of her subjects. A verdict "wrong" was as likely to be followed by the avoidance of a particular choice, as a verdict "right" was to be followed by the repetition of a particular choice.

More interesting was another experiment which at the moment can be described only in general terms. This involved a highly complex activity, depending on accurate visual perception of a moving pattern with many items, the exact location, or judgment, of a particular position within the pattern, and a timed movement requiring the co-ordinate activity of a number of muscle groups. Both experimenter and subject knew what it was desired to achieve, or learn. When the subject had done his best the experimenter knew how near he had come to a perfect performance. Neither knew just how the perfect performance could be achieved. There was a strong motive to learn, for in this case the success achieved meant life or death to a lot of people, if it could be transferred to a different but very similar situation outside the laboratory.

We took two sufficiently large groups, as equal as possible in all relevant respects. One we trained with persistent encouragement

(or reward) and the other with persistent discouragement (or punishment). Both improved significantly at almost the same rate and to almost the same degree within the limits of the learning period assigned. And when the end was reached nobody knew by what precise sequence of processes it had been reached. Certainly it was not reached by going along any fixed chain of links with progressively strengthened or weakened connections. The final trials were with patterns which had never been seen before, and could not be dealt with simply by unrolling an established sequence. But they were dealt with by both groups, with a success that would have been totally impossible at the beginning of learning.

In fact, except perhaps in carefully selected cases, there seems to be something radically wrong with any theory which demands that the acquisition of skill must consist in establishing and strengthening links either between whatever is treated as a single situation and a single response, or between various items in the situation and a series of responses having a definite order. It is necessary to bring in order of sequence because if that is altered the links cannot possibly be the same links ; and this is clearest of all in a view such as Thorndike's, according to which responses are fundamentally movements of some kind. Both these cases, however, the case of fixed single reaction, and that of fixed serial reactions, are uncommon except in specifically arranged experiment.

In normal times a man has two or three great-coats. Having learned to button up one of them he does not, when he puts on another for the first few times, have to learn in the same way to button that, though the buttons and the button-holes may vary widely in size, spacing, position and other respects. He has mastered the buttoning process, and will achieve it readily and with a minimum of conscious effort even though the movements and their connections differ from coat to coat. Similarly, in the second of the two experiments described above, the terminal response (usually considered to be particularly important in a connectionist theory) consists in pressing a morse key. But this can be done just as well by any finger of either hand, and what is often noticed is an inverse relationship between the frequency with which a particular movement has been used and the probability that that particular movement will immediately recur. Indeed, the very great difficulty of transferring a fixed series of responses from one situation to any like situation, which every experimental psychologist has come across, ought to make anybody suspect the adequacy of a simple connectionist theory.

I often think that those of us who live mainly in the psychological laboratory—as I do—would gain greatly if we sometimes relaxed into the arm-chair, and took a little more notice of the works of those who have been much less tied than we are to the limits of carefully invented situations. In learning, for example, we might well reflect again about Lloyd Morgan's notion of " persistency with varied

effort " and the use made of it by G. F. Stout. It is true that both of them look upon the " varied effort " as particularly incidental to a stage of learning, and give less thought to the variation which remains a prominent characteristic of most forms of mastered skill. But if, with this principle in mind, we examine the daily practice of skills, whether of movements or of words, or of any other medium, we shall find that in mastery also, in the majority of instances, it is " varied effort " that more than anything else calls for explanation.

There seems nothing for it but to admit that in some way or another the details of whatever media are used in any form of acquired skill, including remembering, are organised into groups and have interconnections which may vary widely while yet the groups as a whole retain an identical functional significance. The man learns *what* to do and not strictly, except perhaps in very early stages or in specifically routinized cases, any one way of doing it. The link that is of real importance is the link between the situation and its predominant immediate demands, whether these demands are the expression of genes or groups of genes with differing potency, or are the objectives of what Thorndike calls " intellectual powers ". This link can be strengthened or weakened, not only by reward or punishment, in the ordinary sense of these words, but also by many other determining conditions which it is the business of the psychologist to distinguish and study. As the steps towards attempted satisfaction of the demand are taken, various items from the organised groups of media that must be used, appear one after another. They vary freely in form, in number, in order and in qualitative character. Thus both the items which are connected and the sequence of connections may be different from one practice effort to another, and still more from one exercise of acquired skill to another of the same skill. So long as what were called the " demands of the situation " are met, the exact ways in which they are met are usually only very loosely controlled.

Is it possible to give any generalised account of why the same skill will go first along one pathway and then along another ? I think it is, and I think that when Thorndike gets away from his experimental lists, with their limited and defined possibilities of " right " and " wrong ", he indicates one of the leading clues. He says that genes, and gene groups, and their derivative " powers " must be regarded as having differing degrees of potency in any concrete situation. Presumably these relative potencies are inherent in the gene-determined organisation of the individual, but those of the derivative powers simply cannot be regarded as inherent in the same way. In this book, however, they are described as simply present and operating ; there is nothing to show how they have been established, how they are related to the material environmental conditions in any case, and how, when they are active, their simultaneous operation, with differing temporary balance from case to case, helps to explain why the same skill takes now one course and now another.

Thorndike claims to use one other principle only. He does this when he is considering the understanding and utilisation of meanings as they have been expressed in continuous prose passages. He then says that besides " the principle of potencies or weights ", we must use the principle that " the elements must be put and left in the right relations ". Unfortunately the reader is left to put his own interpretation upon this, and given no clue beyond what he can gather from the context. From this it seems as if Thorndike considers that the same elements must be evoked in the same serial order as were present some time or, more likely, a large number of times before.

I suspect, however, that this second principle is a rather obscure way of stating whatever is involved in the notion of " belongingness ", used in this book several times, but elaborated in other writings of our author. Now whatever else is true about " belongingness " it seems plain that it has the character of an abstract notion, and requires a proposition for its expression. It cannot therefore, as I have said already, be a part of man's endowment. The whole drift of Thorndike's discussion suggests that " belongingness " requires simply the reappearance of items and a sequence that have appeared before. In a very precise and literal sense perhaps this never occurs in the whole field of the phenomena of life. But we come nearest to it in the various forms of tropistic and reflex action and in rote recapitulation, the first being usually regarded as a part of endowment and the second as the simplest of all possible forms of learning. To take what may approximately happen at a very lowly level of behaviour as a sufficient model for all levels of action is, to my mind, a lamentable and fundamental mistake.

I have considered the theory of learning which is presented in this book at such length because it is vital for the greater part of the work, and also because Thorndike has, as I think, done better and more experiments on this topic than any other living psychologist. He may have over-simplified his explanation. He may appear to have only the three principles of reward, together with (in some cases) relative potencies, and the maintenance of " right relations ", merely because he has squashed into these three others which are in fact incompatible with them. He may have overdone the fixed sequential characteristic of acquired skill. More than anybody else he has certainly proved that there is something which happens at the culmination of a series of responses which, working back upon the responses and their relations, makes it likely that what the responses achieve will be achieved with greater or less ease in the future. He has also demonstrated in several directions how and within what limits the retro-active influence works. It is, for example, still believed by many that a response to certain classes of situations may be left to become thoroughly established and then suddenly prevented by some drastic and dramatic punishment. Thorndike has shown that it is not so, and that a form of behaviour is not rendered more

likely directly in proportion to the size of the reward or less likely directly in proportion to the severity of its punishment.

I should like to consider in detail the two brilliant, but, I think, theoretically unsatisfying chapters on the psychology of language. I should also like to linger over his discussions of Human Relations, Rulers and Ruled, and Laws and the Law. But the review is already very long and I will be content with a few remarks about the original and interesting contribution which this book makes to the Psychology of Welfare.

There are two main problems: the Welfare of Individuals and the Welfare of Communities. Both are approached in Thorndike's characteristic list-making manner. I propose to consider the second only.

Thorndike draws up "A bill of specification of a good life for a community". This contains seventy-two items. They all refer to life in the modern American city. The first thirty-seven items are crucial and their combined "score" constitutes what Thorndike calls the "G index of the general goodness of life in a city". The remaining items yield subsidiary information which, treated by a variety of statistical methods, may, Thorndike thinks, throw at least some light on what are the main conditions of social welfare. All of the first thirty-seven items can be given an objective expression. They are such features as "Infant death-rate reversed"; "Average salary of an elementary school teacher"; "*Per capita* public property minus public debt". A good many of the subsidiary items are difficult to measure, and some, like "Justice of voters to tax-payers", or "Intelligent sympathy of old with young", invite opinions merely.

Thorndike claims no finality for his list, and recommends anybody who thinks he can do better to try. It is certain that most people of his own level of culture and with his social background will accept the enumeration as on the whole just, and probably sufficient.

He now took 295 American cities of from 30,000 to 500,000 population and got the facts about them for all of the measurable items on his "bill", and opinions for all of the unmeasurable items. Any city that was as low in every item as the city that was lowest of the 295 would gain a G score of 0 and any city that was as high in each item as the city that was highest in the 295 in that item would gain a G score of 1541. The actual range in the case of these particular cities in the year 1930 was from about 300 to about 1100.

It is not possible here to consider the statistical methods used by Thorndike in his attempt to analyse his G score and to discover the conditions which make it high or low. Some of them may involve rather dubious assumptions about the distribution of the characters measured. Thorndike has, however, few peers in this field, and on the whole the work is meticulously careful as well as ingenious. The conclusion is that two groups of factors are more important than any others: the *per capita* income of a city's residents, and their

level of intelligence, character and ideals. There is, naturally, something common to these two and there is an unnamed residue, probably, though Thorndike does not say so, specifically social in its character.

More interesting perhaps are the suggested negative conclusions. Apparently the welfare of a community, as measured in this way, has little or nothing to do with its size, and nothing at all with its wealth except in the form of the highest level of income for the greatest relative number of persons. The G score is not necessarily depressed in communities which have developed large scale mass production, or raised in those which maintain a large Church membership. There is no evidence that the political system, or form of government preferred, is of any particular importance.

Criticism is very easy. It will doubtless be said that the whole approach is superficial; that the "bill of specification" is highly selected and, in any case, can have regional and temporal significance only; that in the present state of record keeping even the most sharply defined of the measures used are bound to be highly approximate; that no amount of statistical manipulation can transform inaccurate data into accurate conclusions; that at best all that we are given are certain current and conventional standards of social welfare, together with a rough approximation to the degree in which these standards are attained in more or less comparable groups.

All these objections are valid. In particular Thorndike does often write as if he is making a genuine contribution to the ethics of social life, whereas all that he shows is that if certain social ideals are good, then these are likely to be approached in fact when certain aspects of training fundamental to them are vigorously developed, and certain others are given less than their common current consideration. This seems to me a great step forward. It means that when social welfare is being discussed, within the limits of the scheme we do know reasonably definitely what we are talking about, and it means that a way has been found of comparing communities which makes it possible, however tentatively and approximately, to range them in an order of achievement. Moreover, again within the limits of the method proposed, it indicates what can be done, especially by way of improving records, to make the ranking more certain and reliable. Between sixty and seventy years ago the first tentative efforts to measure individual intelligence level could have been criticised in much the same way as these proposals can be attacked now. It may be that Thorndike's courageous attempt to deal with the much more complex problem of the level of social welfare may have as great and as beneficial a development as has already been secured in the field of intelligence.

I began to read this book with a very lively regard for Professor Thorndike and his works. He has a long and enviable record of ingenious, simple, unconventional and well directed experimental achievement. There is, I fear, no getting away from the fact that

many of his earlier publications, though they nearly always describe genuinely original work of very high quality, are over-long, wordy, detailed, and contain drawn-out dull passages. There are none of these here. He is summing up the main results of a life-time of research. He does it with remarkable economy and outstanding clarity. So I should like to set it down that I finished my reading with my regard for the author enhanced and my admiration raised to a higher level. It seemed to me to matter little that I could not agree with all his arguments. I knew what he was arguing about, and the whole work is written with that perfect honesty, simplicity and directness which are the marks of a master.

F. C. BARTLETT.

Introduction to Semantics. By RUDOLF CARNAP. (Studies in Semantics, Vol. I.) Cambridge, Mass., Harvard University Press; London, H. Milford, 1942. Pp. xii + 263. $3.50, 20s.

Formalization of Logic. By RUDOLF CARNAP. (Studies in Semantics, Vol. II.) Cambridge, Mass., Harvard University Press; London, H. Milford, 1943. Pp. xviii + 159. $3.00, 16s. 6d.

THESE books are the first in a series of studies planned by Professor Carnap. They deal with difficult and controversial questions whose complexity even Professor Carnap's lucid exposition cannot much diminish. Readers who are quite unfamiliar with the aims and methods of "semantics" (that is to say, the systematic theory of the relations between signs and the entities to which they refer) will do well to look first at Carnap's more elementary survey in the monograph *Foundations of Logic and Mathematics.*[1] More advanced students will need to refresh their memories of the same author's previous work on *Logical Syntax of Language,*[2] a knowledge of the doctrines in which is indispensable to full understanding of the more recent developments.

The general purpose of the earlier book was to demonstrate the potentialities and philosophical relevance of the *formal* study of "languages". A "language" is, for this purpose, understood to be a calculus, a system of uninterpreted symbols of various recognizable kinds, arranged into series by means of explicit rules of

[1] Vol. i, No. 3 of the *International Encyclopedia of Unified Science* (University of Chicago Press, 1939). Chs. i and ii are especially helpful. Professor Alfred Tarski's recent paper, "The Semantic Conception of Truth and the Foundations of Semantics" (*Philosophy and Phenomenological Research,* March, 1944, pp. 341-376) is another excellent elementary exposition.

[2] A useful analysis of the German edition (1934) will be found in Susan Stebbing's critical notice (this journal, Oct., 1935, pp. 499-511). The English translation, with revisions and additions, was published in 1937.

formation and transformation. The rules defining the permissible modes of combination and replacement are called *formal*, because they refer only to the kinds and serial order of symbols and not to their "meaning" or "sense": *syntax* (or *logical syntax*) is the orderly elaboration of such rules and their deductive consequences.

The precise definition of syntactical terms and procedures made possible for the first time the exact formulation of issues of considerable importance in pure and applied logic (among which may be mentioned the definition of 'analytical' or logically true, and the possibility of so interpreting a language as to make it contain a formulation of its *own* syntax). The merits of Carnap's contributions to these and kindred topics have been fully recognized by specialists in the methodology of the deductive sciences; but philosophers whose major interests have little relation to the technical problems of formal logic have naturally shown more interest in Carnap's thesis (also advocated by other members of the school of logical positivism) that the main problems of analytical philosophy *could be replaced* by questions formulated in syntactical terms and capable of solution by syntactical methods. On this question the consensus of opinion has been that the equation of analytical philosophy to logical syntax, though ingeniously argued, was far from proven; and that the weaknesses of the argument might be due in part at least to a certain poverty in the technical equipment of "logical syntax". There is considerable interest, therefore, in noting the modifications of standpoint which five years of study and reflection have produced.

It may be said at once that the changes are extensive. In the earlier researches the characteristic tone was that of a tough-minded determination to exclude all philosophical questions which were not capable of translation into syntactical formulations; thus ostensible references to the denotation or connotation of terms were regarded dyslogistically as belonging to the "material mode of speech" whose use leads to "deception and self-deception"; and terms such as 'universal' and 'proposition' (or even 'number' and 'relation') by virtue of reference to alleged non-formal characters of languages merited opprobrium, and relegation, by all sober thinkers, to a new index of philosophically improper vocabulary.

It comes as something of a shock, therefore, to find these new studies re-admitting wholesale those ostensible referents of symbolism which had seemed to have been permanently banished from "scientific philosophy". It is advisable, we are now told, to investigate the relations between a language and the objects to which the symbols refer (the so-called "designata"); indeed without such investigation it is impossible to define 'truth' and 'falsity' or properly to represent the necessary truths in which these or equivalent terms occur.

The central notion of the new discipline is the relation of *designation*, holding between a symbol (or expression) and the entity which is its meaning. The systematic construction of rules defining the

designation of expressions in a language, together with the determination (in the form of theorems) of the consequences of those rules, provides the *semantics* of that language. As in the *Logical Syntax* Carnap's purpose was in large part restricted to tracing the deductive consequences of his own principles, so also in the books under review much of the exposition is as purely deductive as projective geometry. But there are scattered references to questions of more general philosophical interest and a somewhat abrupt modification of the earlier thesis of the identity of analytical philosophy and syntax. Whatever can be salvaged from the non-empirical portions of the studies now labelled 'philosophy' will prove to belong, we are now assured, *either* to syntax *or* to semantics (p. 250). But no reasons are adduced and we are left to suppose that some modification of the arguments used in the *Logical Syntax* will suffice.

It is too soon to know whether the methods of semantics, in the form in which they are presented by Carnap, will prove serviceable even for the restricted purposes of the methodology of interpreted deductive systems. The presentation of the central notion of *designation*, as Professor Alonzo Church has pointed out,[1] would seem to involve some confusion whose elimination may entail considerable modification in the theory. And apart from such specific difficulties, one has an impression that the increasing complexities of the present system (as compared even with the *Logical Syntax*) are yielding diminishing returns. In syntax we make shift with two languages (the *metalanguage* in which the syntax is formulated and the *object-language* to which reference is made); in semantics, we seem to need an indefinite number; and the need to discriminate symbols according to their position in one or other of these segregated but related systems produces a maze of symbolism which few can tread with confidence.

Perhaps familiarity will bring mastery of even these complexities; and a younger generation of philosophers (if that is the proper word) may find the correct use of quotation marks and gothic letters as easy as epistemology and a great deal less disturbing. But meanwhile it may be permissible to register some qualms and reservations.

The vocabulary of semantics (as it is constructed by Carnap) not only readmits such terms as 'proposition' and 'property' but permits reference to the designata of *any* kind of symbol. ("It is convenient to adopt for semantical discussions a use of the term 'designation' which is wider than the ordinary use, so that we may

[1] See *The Philosophical Review*, May, 1943, pp. 298-304. Church uses Carnap's semantical terminology to reproduce an argument originally used by Frege to show that a sentence designates (bedeutet) a truth-value. Church arrives at the thoroughly unwelcome conclusion that in Carnap's treatment "the designata of sentences . . . must be truth-values rather than propositions". A solution of this difficulty will probably require the discrimination (and appropriate codification) of *several* senses of meaning.

speak of the designata not only of individual constants and predicates, but also of functors and sentences", p. 49).

Here a certain type of reader will irresistibly be tempted to risk the disapproval of the pure semanticist by asking questions of an old-fashioned philosophical sort. He will wonder whether such a multiplicity of types of designated entities is necessary or even "convenient"; and whether some risk of "deception and self-deception" is not re-introduced? Perhaps questions as vague as these (no doubt incapable of exact formulation in a formal system) are symptomatic of incorrigible tendencies to metaphysical regression and deserve no better answer than a contemptuous silence; but they will inevitably recur. They will become insistent, for example, in connection with the so-called "absolute concepts" (*I.S.* p. 88: "We decide to apply some radical terms, *e.g.* 'true' not only to expressions, *e.g.*, 'the sentence 'P(a)' is true in S' but also to the designata of those expressions, *e.g.* 'the proposition P(a) is true'. *In the second case, no reference to a language system is made; the concept is not a semantical but an absolute concept.*" Italics inserted.) Similar doubts arise in connection with the alleged nonconventionality of logic ("... the construction of a system of logic ... within a given semantical system is not a matter of mere convention; here the choice is essentially limited if the concepts are to be adequate ..." p. 247). The reference to the "adequacy" of the concepts introduced does not help much, for it turns out that a definition is adequate if it conforms to a certain requirement stated in a metalanguage. But how do we tell whether the requirement of adequacy is itself adequate—by retreat to yet another language or, at some point or other, by some appeal to current usage?

The philosophical qualms expressed above have a single origin in uneasiness concerning the viability of the abstractions upon which Carnap's semantics is based. The basic symbol situation is regarded by him (following Professor Charles Morris) as involving a *three-termed* relation between a speaker (or hearer), a symbol, and an entity to which reference is made (a "designatum"). But it is by no means obvious that this is a tenable analysis of the factors involved in the declarative use of language; to treat symbols as names for "designata" may be indeed to re-introduce the misleading hypostatizations from which the theory of descriptions delivered logic.

Consider an analogous procedure in the economic theory of prices. Suppose we chose to say that the basic situation involved a three-termed relation between a buyer, a monetary value or price, and a commodity. Let us call the price the *market-cost* of the commodity of which it is the price. If we should now proceed to develop a systematic deductive theory of market-cost (and notions definable by its means) in abstraction from any reference to actual buyers or indeed to any concretely exemplified economic data, and without further analysis of the primitive relation, we should have a close analogue to the procedure adopted in pure semantics. Now it is

XUM

true that economics is an empirical science and pure semantics, as Carnap understands it, is not; but the analogy is sufficiently close to be disturbing. A " pure " theory of market-costs might prove to have some useful empirical interpretation; but most economists would view with justified suspicion any claim that such a procedure could demonstrate the presence of " absolute " relations between commodities (relations invariant with respect to all systems of prices); and a similar degree of caution would seem proper in respect of the parallel claim by Carnap in respect of relations between *designata*. The merits and limitations of semantical methods as philosophical instruments will become clear only when more systematic attempts have been made to apply the general methods to the investigation of " historically given " languages; and we need considerably more light upon the assumptions involved in making such concrete application.

These questions would not be pertinent if pure semantics were, in fact, what it is at times presented as being—a branch of pure mathematics with no specifically intended and preferred interpretation. But of course the pure mathematics of the system is cultivated for a purpose—and a purpose having implications outside formal logic (as appears for instance in the remarks made in passing on ' logic of meaning '—" I now regard semantics as the fulfilment of the old search for a logic of meaning, which had not been fulfilled before in any precise and satisfactory way ", p. 249). Those who are interested in such implications will be disappointed to find philosophical questions treated perfunctorily in the *Introduction to Semantics*, and will hope that Professor Carnap may at some future time set out extensively and in detail his present conception of the epistemological foundations of semantics and of the relation of the subject to general philosophy.

The critical tone of the foregoing remarks should not be allowed to obscure the genuine excellences of this introduction to semantics. Considered independently of its philosophical import (and so, as an essay in a certain type of pure mathematics), the presentation ranks high in lucidity and clarity (qualities which have always been prominent in Carnap's writing); though the work is modestly presented as " only a first attempt ", it is certainly a great deal more than such an understatement might suggest to the uninitiated.

The book entitled *Formalization of Logic* applies semantical methods to the study of the propositional calculus and the logic of functions. Suppose that the propositional calculus formulated as an uninterpreted system contains as connectives only the signs ' $\sim$ ' and ' $\vee$ '; next suppose that a *true* interpretation be given (by means of semantic rules formulated in an appropriate metalanguage), *i.e.* one which translates all the axioms and all the consequences of the axioms into true propositions: how closely will any such interpretation conform to our intuitive notions concerning negation and disjunction? In particular will it be the case in every such

interpretation of the propositional calculus (*a*) that if ' p ' is true in the interpretation, then always ' $\sim$ p ' is false, (*b*) that if ' p ' is false in the interpretation, then always ' $\sim$ p ' is true? Carnap is able to establish the somewhat surprising fact that the answer must be in the negative, *i.e.* that " true " interpretations of the propositional calculus may fail to express our pre-analytical knowledge of the meaning of negation and disjunction. This circumstance described as the existence of " non-normal " interpretations, is here regarded as a defect of the customary systematic presentation of elementary logic; Carnap therefore seeks, and claims to have achieved, a " full formalization " of the propositional calculus. But the notion of " disjunctives " which he uses for recasting the calculus appears to involve a covert introduction of semantical notions and so to beg the question of ensuring full formalization by exclusively *syntactical* rules.[1] It seems indeed plausible that the exclusion of " non-normal " interpretations of a calculus will always involve recourse to *semantical* rules. And it still remains to be determined whether appeal to semantical rules is metaphysically neutral and does not rather involve at some point the making of philosophical decisions concerning philosophical issues.

MAX BLACK.

Happiness, Freedom, and God. By C. A. RICHARDSON, M.A. London: G. G. Harrap & Co. Ltd., 1944. Pp. 221. 7s. 6d.

IN 1920 Mr. Richardson published his *Spiritual Pluralism* and developed there a metaphysical theory of reality as a community of spirits. In this new volume he takes substantially the same ground, but he amplifies the statement of his position by outlining his ethical views and discussing the unity manifested in the plurality which it was the main purpose of his earlier book to describe. He has taken great pains to make himself clear; he writes limpid English and ends each chapter with a useful summary; if despite these aids I still sometimes fail to understand, the fault is surely mine. The great merit of his book is the earnestness and thoroughness with which he develops his monadism. He shirks no difficulty and has the courage to express his convictions at a time when speculative metaphysics is not universally popular. To me at any rate his argument is always interesting, even if at the last I remain unconvinced. In a short summary it is hardly possible to do justice to his case or to the wealth of his material, but I must first give some brief indication of the book's main trend.

Each of us is aware of himself as a unitary experiencing subject; this subject has a structure but, as an organic unity, it is not to be analysed away into anything less than itself. The awareness in

[1] On this point see Alonzo Church in *The Philosophic Review*, Sept. 1944, p. 496.

question is not knowledge: the ego (as Kant says) stands in its own light and therefore cannot be an object of knowledge; but it can be "realised" as the subject, and never the object, of experience. Since the subject realises itself as a subject, it has being-for-itself and therefore also possesses being-in-itself; this means that it is a substance, and indeed the only substance. Reality consists of a plurality of these substances or spirits. The object of our experience is the mode in which the activity of other spirits is manifested to us.

All spirits are active, and indeed experience is simply activity. But the nature of the activity varies with the variations in degree between the spirits. The spirits whose activity is manifested to us as a stone are non-personal; persons are active in the quest for happiness, and while most of their activities are means to that end, there are two forms of inter-personal activity which give rise to happiness in and by themselves, namely (*a*) creative activity in the intellectual and æsthetic fields and (*b*) the active development of satisfying personal relationships. The creation of patterns and relationships of this kind is the "ultimate state" of subjective being.

Reality is an organic whole, and its unity depends on God who transcends all other spirits but is also immanent in them. While the traditional arguments for his existence are rejected, reasons are advanced for believing in his existence as a person who knows, wills, and feels in ways analogous to, though infinitely higher in degree than, our own. God's purpose is the perfect happiness of all, and this purpose is necessarily realised through his immanence in all other spirits. The good is happiness, or what brings happiness, while evil is the converse; and evil exists of necessity because perfect happiness is the end, and happiness at any level is made greater if consciousness of unhappiness has preceded. If we are wise, we shall co-operate with the purpose of God, who knows what will bring happiness, but we must not say that we "ought" to obey God's will or to seek happiness, because "ought" implies alternatives between which we can choose, while we do in fact always seek happiness alone, and our actions are all determined, partly by the actions of other subjects and partly by our own nature. Moral responsibility is a meaningless idea, and freedom means only that each subject is to some extent unique, so that his actions cannot be rigidly forecast, although human action as a whole accords with statistical laws. Non-personal subjects, though unique to some extent, have less uniqueness than persons have, and this is why physics can forecast their movements more accurately than the actions of persons, but still not with perfect accuracy, because the laws of physics are also statistical.

It might seem from what has been said that God is carrying out his purpose through time and that the development from non-personal to personal subjects has taken place through time. But this would be an error. The unitary subject is not in time; time is in it. When we say "duration", we only mean "existence"; and if we describe

a development in serial terms all we are doing is analysing an experience which is not in time at all. If different maps of a whole region are projected from a single point, the points on any given map are in latitude and longitude, but the maps themselves are not; so, too, events analysed in your experience and events analysed in mine may all be in time, but you and I are not. Hence we " may not say that a monad progresses "; " progress is an attribute of the unitary experience which is the monad ". (I quote this but I do not understand it.)

Mr. Richardson ascribes some importance to his theory of causality and time. The causal law which he advocates runs as follows: " A ' state ' is that which *may* occur to an observer as an event. If a state x occurs, there are principles which enable us to determine, at least theoretically, the probability of any other specified state "; this then leads to a conception of " true time " as that " in which the ' past ' and ' future ' of a given event are not other events but the field of the probability values of the event in question ". The author intends this theory to apply to all fields of experience, but in the longest chapter of his book he works it out in detail with special reference to physics. Of this chapter, which he admits is out of key with the rest of the book, I shall only say that its mathematical argument is beyond me.

As Mr. Richardson admits, his case rests on certain fundamental assumptions and on the acceptance of principles which he calls " ultimate ". I am often perplexed to know why he finds his assumptions and principles plausible, but instead of criticising them, which might be unprofitable, since he says that they are " beyond argument ", I will confine myself to discussing one or two points at which I find his position hard to understand.

(*a*) My first difficulty springs from an inability to reconcile his views on time with his insistence on " progress ", the " ultimate " realisation of God's purpose, and the " development " of the monads. I could understand him better if development in time were regarded as an appearance, or even as a serial analysis, of timeless categories or forms whose contemplation is the eternal activity of God and the other monads. For example, before there can be human life there must be a rhythmical movement of blood to and from the heart, and this movement takes time; but it might be held that this is only an exemplification, in terms of that conceptual construction which we call space-time, of a timeless concept of life, just as the three types of triangle which we can draw or imagine merely exemplify characteristics analysable in the concept of triangle as such, a concept which can neither be drawn nor imaged. It might then be held that the truth of objective reality consisted in a system of timeless concepts, and that for an infinite mind which could think them simultaneously the universe would be timeless, although finite minds would only be able to comprehend them one after another and so would necessarily think serially and import time into their

XUM

conception of reality. Occasionally Mr. Richardson seems to approach this point of view, for he tells us that the structure of experience can be reduced to probability laws statable in mathematical equations, and that these can be grasped by God simultaneously, while finite minds have to set to work analytically and construct descriptions of experience piecemeal and serially. But one difficulty here is to reconcile this with Mr. Richardson's conception of monadic activity; for this activity is not confined to an eternal contemplation, whether of mathematical equations or of æsthetic patterns; it also consists in the "development" of personal relations, and such a development seems to imply time in a way in which eternal contemplation may not. A development is a synthesis of permanence and change, and Mr. Richardson specifically says that such a synthesis is implied by his metaphysics; and yet time would seem to be not so much an abstraction from change (as it may be for him) as one of its presuppositions. Again, we are told that in the monadic activity of which reality consists there is a "process" in which those manifestations of monadic activity which we call the material world and evil are dying away (in accordance with the Second Law of Thermodynamics) as monads progress from a rudimentary level to the level of spontaneity; and this again seems to me to presuppose time as that during which the monads develop. The view that time is not an ultimate reality but was created by God as a moving image of eternity is of respectable antiquity; and if Mr. Richardson were merely insisting on something like this, his view would be less paradoxical. But he rejects the idea of Creation; God and all the monads are apparently existent from everlasting to everlasting; but if so, then process disappears out of reality, and the changes in monadic activity which are to result in pain's passing away are mere appearance or illusion. And I do not know what then becomes of the personal relationships to which Mr. Richardson attaches so much importance. He may reply that what he has in mind is the harmony between all souls who enjoy the presence of God in heaven. If so, I am reminded of Sir Thomas Browne: "If any have been so happy as to understand Christian annihilation, . . . they have already had an handsome anticipation of heaven; the glory of the world is surely over and the earth in ashes unto them." For myself, I cannot, so easily as Mr. Richardson, get past "the glory of the world" nor have I been so happy as to be able to explain its history in terms of mathematics. It is true that he says that to transfer to the subject the temporal relations subsisting between parts of the object is unwarrantable; but if my memories are part of the object, as I suppose in his view they are, I wonder what is left of the subject and the happiness it seeks to enjoy.

(*b*) My second difficulty concerns the character of subjective activity. Spirit alone, we are told, is real; matter is a concept constructed from the way in which we experience our interaction with monads at the most rudimentary level. (Why we should

construct this fiction we do not seem to be told.) But although spirit is the sole reality, the activities of the monads vary at different levels, since there is a tendency, visible in the evolution of species, for monads to develop steadily from lower to higher levels of mentality. I take this to mean that when I have experience of a stone, a mouse, or a friend, I am experiencing the varying activities of spirits at varying levels of their mental development. If Mr. Richardson were simply calling attention to a continuity of principle underlying the differences between matter, life, and mind, as seems to be implied by evolutionary science, there would be no unusual difficulty. But theories which trace the evolution of life from matter, and mind from life, ascribe to process a reality which Mr. Richardson seems to deny, and they commonly allow that the difference, *e.g.* between life and matter, is one of kind as well as degree, and this he apparently rejects. He is scornful about the theory of emergence, which in his view explains nothing but only describes what occurs; but it does describe something, namely the occurrence of novelty in evolutionary development, which he seems to explain away by making all differences between monads, even between man and God, differences of degree only. If he took the notion of development as seriously as he takes the notion of spirit, would he not be compelled to recognise differences of kind between his spirits and so to reintroduce matter as a reality, however he might christen it? In this connection he might perhaps have developed further his distinction between non-personal and personal monads, but instead of doing so he insists on the generic similarity of all monads, *e.g.* by holding that all alike are amenable to the same statistical laws. But why should he suppose that the same type of laws applies both in history and in physics, both to immature monads and to those more highly developed spirits which he calls persons? What kind of activity it is which is manifested as a chair we must presumably infer from the chair, or perhaps learn from physics; but in either case the activity in question seems to be so different in kind from that which we experience in forming satisfying personal relationships that to make both kinds of activity amenable to laws of the same type is not so much to clarify our experience as to conjure away its differences. And I am perplexed about Mr. Richardson's meaning, because he does not ignore the differences.

(*c*) Finally, he begins by distinguishing between description, which is the task of science, and explanation, which is expected from philosophy. Ultimately this expectation must be falsified, but to describe the universe in terms of subjects is to explain it in a sense, because we are subjects ourselves and thus, in so describing it, are starting from a realised fact of our experience. The author goes on to warn us against pressing analysis too far; everyday language and ideas often have an importance and a significance which the philosopher must reckon with and not analyse away. This is an encouraging start, but does Mr. Richardson abide by his principle as he

proceeds? We may grant that we "realise" the existence of "subjects" and yet wonder why a description of reality in terms of them alone should really "explain" our experience; is our ordinary consciousness not just as much convinced that matter exists (however it may be described) as that spirit exists? I suspect that though Mr. Richardson has been prepared to swallow large doses of positivism, he has stopped short of accepting an analysis of the self as a bundle of sense-data. But why have his reasons for rejecting that nostrum not been used to make him doubt his initial acceptance of positivism elsewhere? He describes his practical philosophy as a "combination of metaphysical theism and ethical positivism". While we may admire the resoluteness with which (like Kant) he imposes barriers to positivism and empiricism, we may wonder whether his blend of positivism with the belief in the reality of spiritual substances is really plausible, especially when he tells us that ethics is a derivative of metaphysics and logic. For my own part, it is his ethics which I find hardest to accept. In reply to the question: "Why do you do this?" a modern moral consciousness may reply: "Not because it will make me happy, but because it is my duty; the fact that a dutiful action may incidentally make me happy need not be any reason for doing it". Now this may be an error, and duty may be an illusion, but Mr. Richardson is committed to explaining our ordinary experience and language, and to call duty a meaningless conception is not to explain it, still less to account for its existence as a concept. He suggests that saints have been moralising (in terms of duty) for some thousands of years, with what success "may be judged from the present state of the world". But the moralising has not always appealed to the same moral theory, in particular it has not always used the conception of duty. Certain courses of action have often been recommended as leading to happiness in heaven or discouraged as leading to misery in hell. And Mr. Richardson's own moral theory, based on happiness, and on wisdom as the means of attaining it, is drawn in essentials from ancient Greece and has thus had a longer time to affect the world than the theories about duty which he rejects. As to the merits of his own theory, I will only remark that he has his own definition of happiness; he implies that all subjects will find their happiness, as conceived by him, in the same activities, and he believes that there is empirical evidence as to what those activities are. Does this evidence really justify his view that what will satisfy everyone, Indians and Africans, scholars and airmen, artisans and artists, is the creation of intellectual and æsthetic patterns and the cultivation of personal relationships? A man may believe this, but is there empirical evidence to prove it to someone who strenuously upholds the uniqueness of individual spirits and the unpredictability of their several actions?

T. M. KNOX.

V.—NEW BOOKS.

The Gospel of the New World. By OLIVER C. QUICK, D.D. London: Nisbet and Co. Ltd., 1944. Pp. xiv + 119. 6s. 6d.

THIS volume contains the last lectures which the late Professor O. C. Quick delivered as Regius Professor of Divinity at Oxford. They were prepared for publication by the author shortly before his death. In an informative preface the late Archbishop of Canterbury pays a moving tribute to the personality of the author, singling out for special praise the clarity and sincerity of his thought. The lectures themselves endeavour to substantiate the main tenets of 'traditional' theology on the basis of a special significance ascribed to the "cosmic eschatology of the New Testament" (p. 66).

Much that is subsidiary to the main theme of the book invites discussion. We must confine ourselves to the latter. This turns very largely on a distinction which is drawn between sin and guilt. The individual is not, as the strict traditionalists are apt to think, "responsible for sins which he could not help committing" (p. 47). Although moral evil, and, indeed, as is also stressed, every positive evil, is due to the misuse of his freedom by the finite creature, wrong choices in the past, on the part of the individual himself or of others, "distort his perception of what is right and weaken his will to do it"; they "have affected for evil the environment in which individuals are born and nurtured" (p. 46). We have thus evidence in our own experience of the fall, a 'fatal choice' by which sin "has gained a foothold in human nature" (p. 39). "Henceforth man's nature is marred" (p. 39). He has "forfeited the immediate constancy of God's help" and must "choose without clear insight to guide his choice" (p. 40). The choice is "a real element in sin" (p. 47) and the power of making it resides, as we read with some relief in view of the main trend of Protestant theology to-day, "in the individual man, not in the group, society, race or nation" (p. 36). But individual choice does not exhaust the whole nature of sin; it accounts for guilt, for the *reatus* but not for the *vitium* of sin. This was obscured in the teaching of the prophets who thought of sin "too exclusively as the immediate choice of the individual will" (p. 46); but a corrective was supplied by the penitential psalms and the priestly religion of post-exilic times. This is how the latter, providing ceremonies of expiation for uncleanness accidentally contracted and for wrongs done in ignorance, helped to prepare for the Christian conception of atonement. The latter is a "mighty act of victory" (p. 89) by which the dominion of sin in the world is broken. It is the giving of new life as well as the forgiveness of past sin. The doctrine of atonement is thus inseparable from the doctrine of the resurrection and the eschatology of the 'new birth', "the life of the world to come" (p. 94). This is for the true Christian the beginning of the "new creation or of eternal life" (p. 60), and it is also a means of the "redemption of the body" and a transformation of the whole material universe by which it is to be "spiritualized and become perfectly responsive to God's Spirit" (p. 62). This found expression especially in the teaching of St. Paul, St. John and the author of Hebrews. After the apostolic age, as we are shown, the emphasis shifted to the idea of individual justice rather than cosmic redemption. This was largely due to the effect of Neo-Platonism. But mediæval legalism also helped to present the problem of sin in terms of merit. Here

XUM

atonement refers to the past only. This juridical view was followed for the most part by Lutheran and Reformed theologians, although the doctrine of salvation by faith brings their teaching "superficially nearer to that of the apostles" (p. 84). We have to recover to-day that which was lost at the time of Constantine. This is the idea of Christ, in his death, not as himself bearing the penalty of sin, but as the sacrificial victim offering in full and voluntary sacrifice a sinless life. An essential paradox of Christian doctrine is this combination of the "extreme opposites of sacrificial and penal death" (p. 101). "Thus the utterly unholy and common death which is the due fate of all men, is itself by Christ's love transformed in his own case into the entirely holy death of perfect sacrifice" (p. 101). This is "the mystery of the blood of Jesus, giving both expiation and new life" (p. 103). The Church, by descending into the world, can, in like manner, "offer itself vicariously in the world's behalf" (p. 105).

The main merit of this theory is the challenge which it sounds to the despair and defeatism so evident in theology to-day. But it is also fraught with considerable difficulties. Among those which will occur most readily to philosophers are the following: (*a*) Is it meaningful to speak of sin other than the sin for which some individual is responsible? Can we not give an account of 'material' wrongdoing in a way that does not presuppose some sharing of sin? This is one of many issues in respect of which a closer contact between the philosopher and the theologian would be very beneficial. Quite often, although dealing with the same problems, the one speaks in terms hardly intelligible to the other. (What, for example, are moral philosophers to make of the Barthian distinction between moral sin and religious sin?) (*b*) Can ethical blindness be wholly ascribed to sin? May not perplexities be due to defective moral sense and ignorance of fact for which no one is responsible? (*c*) How are we to establish the relation of sin to the 'doom' of death under which man lies? (The author seems to make some play here with the metaphorical meaning of 'dying to oneself'.) (*d*) Does not the account of atonement and resurrection seek to weave too close a logical pattern out of ideas whose significance is mainly symbolical? (*e*) How can our physical environment—or the physical body for that matter—be infected with sin, and why should the salvation of "man's whole being" (p. 57) require transformation of the material world? (*f*) Can all evil be reduced to moral evil? It is urged in Lecture I that pain and error are not positive evils except in so far as they are willed. This appears patently untrue. However different from moral evil, pain and error are sufficiently evil in themselves to constitute a very serious problem for any philosophy which has at its centre the idea of a beneficent deity.

It is not easy to form a final estimate of the worth of Professor Quick's book. The reader cannot fail to be impressed by the conciseness and clarity of the style. We find also that mastery in the arrangement of material which is so distinctive of the author's other writings. The scholarly allusions are invariably helpful, and the accounts of great movements of thought in theology not only illumine the main theme of the book, but are of much interest in themselves. But any final judgment of the worth of this kind of book will depend on the importance ascribed to the theological assumptions (and the view of the authority of Scripture) on which it proceeds. It will not have much appeal for readers not in some sympathy with traditional theology.

H. D. LEWIS.

Philosophy. By C. E. M. Joad. (Teach Yourself Books.) London: Hodder & Stoughton Ltd., for the English Universities Press Ltd., 1944. Pp. vi + 228. 3s.

This book belongs to a series which covers many arts and accomplishments, and it is designed for the educated beginner. Professor Joad has written a book with a purpose: to help the distracted peoples, or at any rate the attentive reader, towards the contemplation of a world of eternal values which may serve to regulate their own living. The book takes the form of a summary-outline—the kind of thing which every teacher of philosophy ought to attempt once in a lifetime for his own private use. It describes philosophy and its various branches; summarises Plato in forty pages; outlines the Empirical and Rationalist criticisms of the familiar world; and comes to some conclusions on ethics and politics. The whole shapes itself into a Rationalistic and Idealistic account of the world which the beginner will find lively and intriguing and which he will probably read to the end.

Professor Joad conceives philosophy to be a distinct branch of enquiry with its own subject-matter and with findings that ought to be expressible in clear statements. The subject-matter is a conspectus of the results of the separate sciences and of history, and the reports of ordinary experience. These together provide a picture of the Whole. The philosopher may be said then to piece together the findings of all the sciences, etc., and, as far as he can, to make one picture out of them. When he has the facts before him, he makes inferences from them and interpretations of them: "he is concerned not so much with the facts as with their meaning; not so much with conclusions as with their significance". This "significance" is the *real nature* of things, as opposed to their outward appearances or measurements or behaviour. And the philosopher is sure to discover that the familiar world cannot give us a picture of the real nature of anything: because its objects either have no fixed character of their own, or are "unintelligible".

The Empirical philosophers for instance showed that the familiar world is not "wholly external to and independent of ourselves". The Rationalist philosophers found the objects and processes of the familiar world unintelligible. The philosopher must have something *real*, and if this something is not the familiar world, he is driven to look for it elsewhere. Like Plato he must take the extraordinary step of "postulating another order of reality, at once underlying and explaining the phenomena of the familiar world".

Professor Joad himself looks for this reality in the realm of Ethics. He examines "Naturalism" and decides against it and fixes on the Forms or Ideas of Goodness, Truth, Beauty and Happiness, as real universals belonging to "another order of reality which is immaterial". From this not unfamiliar ground we make the final ascent: The Mind of God is the only possible "home" for these Forms, or at least for the Form of the Good. In this last stage of his argument, Professor Joad appears to repeat that conceptualistic view of universals which he has rejected in his chapter on Plato; and he does not appear to protect himself (as Leibnitz tried to do) against the suspicion that the moral laws (and even the laws of mathematics) are valid simply because the universal Mind happens to prescribe them. However that may be, Professor Joad reaches the conclusion "that morality points to without necessitating God".

From the Height we look down to the waiting multitude below. What application has all this for human life, society and Government? The chapter on politics is characteristically pungent and provocative, and would nowadays be described as individualistic. One would wish to add some account of the changes that now oblige us to build up a positive moral State on the largest scale.

It will be seen that Professor Joad has written the sort of account of philosophy which a conscientious psychologist or zealous theologian might write on his own subject. That is to say, he modestly admits the lack of agreement in the conclusions, but tries to make things look as compact and tidy and *as like physics or chemistry as possible.* He even goes so far as to hint that the obscurity of expression which characterises philosophy is "nothing but bad craftsmanship"; that it is on account of Kant's laziness or incompetence as a writer that "volumes have been written by his commentators and critics with the object of determining not whether what he said was true, but what he meant by what he said". The innocent reader might imagine that philosophy is only accidentally concerned with explaining the meanings of certain difficult terms and expressions, and that once all these are cleared, the task of the philosopher is to apply some independent test for Truth or Falsity.

Such a suggestion is thoroughly injudicious. All the great philosophers have written about meaning and language in a way which shows that they have a special interest in these problems, and even that "How to make our ideas clear" is a chief aim of the philosopher. If so he must be particularly conscious of the pit-falls of language, and especially of all pseudo-technical and emotive language. This conscience, this sensitivity ought to be encouraged in the beginner from the earliest possible moment. Naturally, Professor Joad is himself concerned with questions of language. On page 23 he asks "What is meant by saying of a thing that it exists?" On page 113 he asks what "to be real" means, and on page 20, "What, incidentally, do we mean by 'right'?" But he is inclined to accept and to use certain notorious terms and phrases without any apparent misgivings. For example:—

> "a thing's relations to other things are just as truly *parts* of it as are its qualities" (111);
>
> "there is something *unique* about man's moral consciousness" (162);
>
> the difference between a savage and a saint "may involve *a difference in kind*, as when from the association of two elements, oxygen and hydrogen etc." (136).

It ought to be clearer than it is in this book, that such terms as these must be scrutinised with all one's powers of concentration.

KARL BRITTON.

The Device of Government. By JOHN LAIRD. Cambridge University Press, 1944. Pp. 173. 6s.

THE concept of political obligation may be fitly compared to the Minotaur of Crete. It is a monster that dwells at the heart of a labyrinth of human prejudices and pre-conceptions, but is formidable, and indeed powerfully operative, in the outer air of experience. If men are to live comfortably, someone must go in to deal with it from time to time, and many readers must have been highly delighted when they learned that Dr. Laird

was the next champion in the lists. Nevertheless, a candid reviewer, who knows the author's deservedly high standing as an ethical and political theorist, cannot but express his disappointment that this should be the best defence at present available of the principles upon which British political institutions rest. An essay on civil polity that says nothing about Trial by Jury, Habeas Corpus and Freedom of Speech is surely not in the best British or Anglo-Saxon tradition, and as the writer ignores the liberal developments in Catholic political theory since *Quadragesimo Anno*, he has almost nothing to put in their place.

The most serious defect of Dr. Laird's work is that he makes it appear that the labyrinth is more important than the Minotaur. With the help of all the best previous contestants, like Plato, Aristotle, Hobbes and Locke, he picks his way very skilfully inward, . . . His encounter with the Minotaur seems casual, and he breaks off the contest after a mere skirmish in order to carry back to future contestants an objective account of what he has observed. If, as I believe, he deals the monster a fatal blow, he is quite oblivous of the fact. In any case, his report is unlikely to reach the ordinary man, because he entered without leaving any single clue to guide him back through the mazes of speculation into the realm of actual experience. Indeed, most readers will close the book with an uneasy suspicion that political theory has claimed another not inconsiderable victim for its labyrinth.

Let us begin with that lucky glancing blow, in which Dr. Laird does actually deal with the central topic, and in language that uses the thought-forms of actual experience yet is not inconsistent with the categories of abstract thought. " On the other hand it is not simply negligent to ask whether there is not something brotherly in all human social relationships. Men are born into a social group without being consulted, and, in the main, have to settle down with neighbours not of their own choosing. This is the necessary corrective to political theories, such as the contract theory, which place exaggerated emphasis upon free political association. The ties and loyalties of living and remembering together, generation after generation, are not easily broken " (p. 134).

That surely is the root of the matter. Men are born into families. The heads of families fraternise, sometimes willingly, and sometimes from fear. They alone are, in the first instance, full members of the society so constituted. Their children become members of it either by succession or by finding a master, it may be in their own family, who will concede to them the means of being independent. The story of Esau, Israel and Laban is the story of the re-birth of society in each generation. The device of government is, in the first instance, to entrust the elders with the task of making the decisions, and the juniors with the task of executing them. Out of this implicitly aristocratic scheme, the different forms of state described by Plato in his *Republic* may be developed by steps that are practically possible as well as theoretically conceivable. Yet every step away from it is fraught with the danger of dissociating the device of government from the fabric of society. When that happens the remedy must come from the family, not from the state. Both the English and the Catholic theories are at one on this point. They diverge only in their view as to the means whereby right principles of living are transmitted from generation to generation.

Totalitarianism has attempted to remedy this trouble by regimenting family relationships into the pattern required by the governors, which is

a natural result of letting the juniors do the deliberation as well as the execution. Another contrasted mal-adjustment may be seen in the recent history of the French Republic, where the principle of seniority enervated the administration in spheres like the Army and the Police, where the greatest enterprise was necessary in order to carry out the will of Parliament. The further deduction that conditions may arise under which a society may have to re-constitute itself by restricting full deliberative powers to those who are the effective masters of households may be hateful to equalitarians, but, as Dr. Laird shows, fraternity is the gateway to freedom, not equality.

The most valuable element in *The Device of Government* is symbolised by the title itself. Whatever else Dr. Laird may forget, he shows himself fully conscious throughout that government is a device for the control of society, and that its coercive acts always require public justification. If that prevents governments secretly pursuing policies that are wiser than an electorate can comprehend, the champions of society can only reply that in the long run wisdom is the product of one's general experience of life rather than of specialised activities or studies. The common English belief that specialists are always wrong on questions of ultimate policy, however skilled they may be in executing any given policy, is by no means an idle prejudice. Enthusiasm, not judgment, is the speciality of specialists. The justification for an act of coercion is that it has been publicly sanctioned by men of experience and responsibility within the social group to which they belong, and is being executed by diligent and faithful administrators. When either condition is lacking, the acts of government cannot be validated by society, and the device of government has broken down.

ARTHUR T. SHILLINGLAW.

Verifiability of Value. By RAY LEPLEY: New York: Columbia University Press, and London: Humphrey Milford, Oxford University Press 1944. Pp. ix, 267. 22s.

THE object of this book is to show that value propositions and not only factual propositions are verifiable. This is supported by a great many vague statements bringing out analogies between scientific and ethical judgements which do not seem to the reviewer to cut much ice. No doubt there will be a close analogy as long as we are concerned with instrumental value, and this for the best of reasons, namely, that the knowledge we require for determining instrumental value is simply scientific, because the instrumental value of something as a contribution towards a given end depends entirely on causal laws which can only be established by observation. But the author does not, as far as I can see, say anything which gives strong support to the view that propositions about intrinsic value can be verified in the same sort of way, and his whole discussion indeed seems to me to be confused by neglect to observe this elementary distinction.

However, it is not really his intention to maintain that factual and value judgements are verifiable in the same way, though from some passages one would think it was. His view seems to be rather that every fact is also in another aspect a legitimate object of a value judgement, because it is either an object of interest or at least relevant to an interest, and that these value judgements are verified by finding out whether the object in question does really satisfy desire. At the same time he maintains that all value

statements are translatable into factual statements and *vice versa*. But I do not see how this can either be true in itself or reconcilable with his doctrine that the two are verifiable in different ways. He says[1] that "the scientific statement 'Water is H_2O' can without loss or addition of essential meaning be translated . . . into moral terms such as 'One who would express correctly in this sign system the chemical constituency of water must (or ought to) say that it is H_2O' ".[1] Now the second can surely only be a translation of the first if the moral term "ought" is already present in the meaning of the first, and I do not see how it can be, if only because that statement itself includes no reference to possible volitions. (Some would indeed retort that "ought" is not here being used in a moral sense, but this is not the view of the author, since he talks of translating the first statement into "moral terms", and if it is true it only follows that he has not succeeded in his object of showing that all factual statements can be translated into ethical terms). That verification of value judgements should be conceived in terms of satisfaction of interests is at any rate the least unplausible naturalist view, but the naturalist who wishes to give an account of the verifiability of value ought to work it out in detail with concrete instances as illustrations if he wishes really to clarify his position. This our author does not do. Nor is there much indication in the book that he has really faced the issue between naturalism and non-naturalism, and on the allied issue between a subjective and objective view of ethics he sometimes confuses the view that values are objective with the view that they are outside experience altogether,[2] and is very obscure in his distinction of the different meanings of the terms.[3] It is an extraordinary fact and one that may throw some light on the causes of the book's deficiencies that the words, ought, duty, obligation, right do not occur at all in the index. Without a discussion of such concepts, I do not see how a treatise on ethics can hope to do justice to its topic, and it is just these concepts which in my opinion render the kind of naturalistic approach made fundamentally unsatisfactory. Yet the book leaves me with the impression that the author has something of value to say, and to prevent the would-be reader from being too much influenced by the criticisms of a reviewer, who is after all there primarily to pick holes, I note that Professor Ralph Barton Perry is quoted in the advertisement as saying that the book is "important" and "makes a genuine contribution to the subject".

A. C. EWING.

Received also :—

D. M. Emmet, *The Nature of Metaphysical Thinking*, London, Macmillan & Co. Ltd., 1945, pp. xi + 238, 10s. 6d.

R. G. Collingwood, *The Idea of Nature*, Oxford, Clarendon Press, 1945, pp. viii + 183, 15s.

R. F. Pickard, *Time, Number and the Atom*, London, Williams & Norgate Ltd., 1945, pp. vii + 92, 8s. 6d.

J. Maritain, *The Dream of Descartes, together with some other Essays*, trans. by M. L. Andison, New York, Philosophical Library, 1944, pp. 220, $3.00.

M. R. Cohen, *A Preface to Logic*, New York, H. Holt & Co., 1944, pp. xi + 209, $2.50.

[1] P. 222. [2] P. 214. [3] Pp. 192-5.

L. Hodgson, *Theology in an Age of Science* (Inaugural Lecture at Oxford), Oxford, Clarendon Press, 1944, pp. 16, 2s.

R. Niebuhr, *The Children of Light and the Children of Darkness : A Vindication of Democracy and a Critique of its traditional Defenders*, London, Nisbet & Co., Ltd., 1945, pp. viii + 128, 7s. 6d.

S. Radhakrishnan, *Education, Politics and War*, Poona 4, International Book Service, 1944, pp. 208, Rs. 5, 8s. 6d.

B. Malinowski, *A Scientific Theory of Culture, and other Essays*, Chapel Hill, University of N. Carolina Press ; London, H. Milford, 1944, pp. ix + 228, $3.00, 18s. 6d.

W. M. Burbury, E. M. Balint and B. J. Yapp, *An Introduction to Child Guidance*, London, Macmillan & Co. Ltd., 1945, pp. viii + 199, 7s. 6d.

H. Cherniss, *Aristotle's Criticism of Plato and the Academy, Vol. I*, Baltimore, Johns Hopkins Press ; London, H. Milford, 1944, pp. xxvi + 610, $5.00, 33s. 6d.

The Meditations of the Emperor Marcus Aurelius, edited with Translation and Commentary by A. S. L. Farquharson, 2 vols., Oxford, Clarendon Press, 1944, pp. xi + 431 ; 432-935, 42s.

S. Radhakrishnan, *India and China*, Bombay, Hind Kitabs, 1944, pp. vii + 168, Rs. 6.

A. S. Nash, *The University and the Modern World*, London, S. C. M. Press Ltd., 1945, pp. 223, 12s. 6d.

C. Garnet, *Victory Vision*, Canterbury, J. A. Jennings Ltd., 1944, pp. 39, 1s. 6d.

T. F. Taylor, *Don't Smoke !*, London, National Society of Non-Smokers, 1944, pp. 48, 9d.

J. R. Baker, *The Discovery of the Uses of Colouring Agents in Biological Micro-technique* (Monographs of the Quekett Microscopical Club), London, Williams & Norgate Ltd., 1945, pp. 22, 1s. 6d.

VI.—NOTES.

To the Editor of "Mind".

Sir,

Professor Aaron, in his review of my Berkeley's *Philosophical Commentaries*, points out, as a reviewer is in duty bound to do, that the final verdict on my text must wait till the MS. is once more accessible. But meantime I should like him and other Berkeley scholars to know that the highest standard of accuracy has been aimed at, and that I had grounds for hoping that a reasonable degree of accuracy has been attained, sufficient to warrant publication.

I made my transcript with great care some twelve years ago in summer light when my sight was pretty good. I aimed at taking over with me to Ireland an exact replica of the whole MS. volume. I did it in two separate

notebooks, like the original before binding, because I wanted to study every detail and test for myself the Lorenz transposition which *at the outset*, I did not believe. I spent three weeks, I think, on the transcribing, and on my next visit to London I checked my transcript with the MS. *Homo sum*, and no doubt I have made mistakes; but grave errors would almost certainly have come to my notice in the following years when I was studying the transcript microscopically and comparing my text with the texts of the other editors. Two pages indeed I can check; for I happen to have photographs of the originals (one is reproduced as frontispiece); both pages are correct. I have compared the whole of my printed text with my transcript; a dot is missing from one "i", but otherwise, they correspond, I think, exactly.

When my proofs arrived, I did my level best, with the kind help of an influential friend, to secure access to the MS. for a final check-up. The authorities were sympathetic, but I gathered that my application was just too late; at any rate it was refused. Had I been a younger man, I might have held my hand, but *labuntur anni*, and at the time the decision to print was taken, there was no guessing how long one might have to wait till the MS. was available. Even now, who can say? It looked like a case of 'now or never'. I hope scholars will approve my decision to go ahead, and that the final verdict when it can be delivered will prove satisfactory.

In view of the state of the MS., described by me on page xiii, a photostat edition would not give what is wanted.

A cheap *diplomatic* edition is hardly practicable, but it is hoped to bring out a cheap edition of the resultant text when things are normal.

May I also say something about the change of title? With due respect to the reviewer, I do not lack the courage of my convictions in this important matter. An established name cannot be changed overnight; a transition period and a transition document are here required. I left the old name in the sub-title, and in one or two places in my Introduction where the context demanded it, and (in the abbreviated form *CPB*) in my notes. If I had not done so, my references would have been unintelligible to many scholars at first. But in a work on Berkeley, now in the press, I use only the new title, and have discarded the old title completely.

Has the reviewer himself made up his mind? He admits that the old title 'is perhaps misleading'; he admits that my argument for the underlying continuity and purpose of the work 'is sound'; he demurs to my title *Philosophical Commentaries*, suggesting *Philosophical Comments* in its place; and he attempts to show that the document (*a*) is a commonplace book, (*b*) 'is no ordinary commonplace book', and (*c*) is a commentary as well.

The matter calls for decision now. A responsible literary journal (and literature is keenly interested in Berkeley) has endorsed my case for the change. Why should philosophy lag behind? Here are two small nameless notebooks, which Fraser read all topsy-turvy, completely failed to understand, and in desperation named *Commonplace Book of Occasional Metaphysical Thoughts* (1871), subsequently abbreviating to *Commonplace Book*. This title has misled two generations of scholars as to the contents and nature of the work. Why should we hand on to posterity this *damnosa hereditas?*

The document is not a commonplace book in the ancient sense of that term, or the modern. I know Locke's *A New Method of a Commonplace*

Book; I have seen in the Egmont Papers the actual commonplace book compiled by Berkeley's friend, Sir John Percival: this document and the commonplace book of those days are as unlike as chalk and cheese. Nowadays the term signifies a collection of odd jottings. *The Times* a few years ago, misled by the old name, described the document as 'Berkeley's scrap-book'. How can this serious examination of immaterialism ever come into its own, as long as it is saddled with that name and treated as a scrap-book? Give a dog a bad name and hang him. In the interests of truth the name must be changed, and now is the time to change it when for the first time a diplomatic edition is available, with, I trust, a reasonably accurate text.

The term *commentaries* is elastic. We have Caesar's *Commentaries* as well as commentaries on the Bible, classical texts, etc. I think Berkeley was here commenting on a written work, but even if it were not so, he was certainly writing notes with a definite aim on a definite theme, and the notebooks which contain those notes may justly be described as commentaries. *Philosophical Commentaries* sounds well, and the title gives the freshman a fair notion of what to expect; but call the work a commonplace book, and the teacher has to spend quite a time in explaining that there is nothing commonplace about it.

I felt reluctant, like all conservative scholars, to make the change at first; it seemed like parting with an old friend; but the old friend has come back to me with added dignity and appeal. Those who will make the effort to make the change will find themselves, I believe, not only satisfied but rewarded.

The reviewer's 'complaints' concern *minutiæ* without any real bearing on the printing of the text or the interpretation. I have dealt with almost all of them—the College societies and the *miscellanea* on page xxvii, the overlapping of the notebooks on page xvii—but I know it is not easy to feel the full force of the argument unless one has replicas of the two notebooks at hand for reference.

I am sorry I did not know, or had forgotten, that Professor Aaron had anticipated my dating of the *Commentaries*. I have referred (p. xiv, *n.*) to the useful work he did on the text, and I would gladly have acknowledged other debts, had I known of them or remembered them, more especially as on the very first occasion of my visiting the British Museum to see the MS., I found him in possession, and he very courteously made it over to me. The fact is that when I set myself years ago to master the document in Johnston's edition, I found in the introductory matter so little known for certain and so much that could be doubted, that I tried to forget everything I had read about it, made my mind a *tabula rasa*, took my transcript of the MS., and worked out the whole thing (structure, date, purpose, and marginal signs) for myself *de novo*.

Yours, etc.

A. A. Luce.

MIND ASSOCIATION: ANNUAL MEETING AND JOINT SESSION WITH THE ARISTOTELIAN SOCIETY.

The Annual Meeting of the Mind Association will be held this year at Bedford College, London, on Friday 13th July, at 5 p.m.

It will be followed by a Joint Session with the Aristotelian Society for which the detailed arrangements will be announced later.

Accommodation will be provided in Bedford College.

The inclusive charge for board and lodging from Friday dinner to Monday breakfast will be £2, or from Friday dinner till Sunday after dinner £1 11s. 6d.

Separate meals: Breakfast, 1s. 6d.; lunch, 2s.; tea, 9d.; dinner, 3s.

There will be a charge of 10s. as a Registration Fee for membership of the Joint Session. The papers will be published by the Aristotelian Society as a Supplementary Volume, which will be sent free of charge to all who have paid the Registration Fee.

Applications for membership and accommodation should be made before 15th June to Prof. H. B. Acton, Bedford College for Women, London, N.W.1. Cheques should be made payable to Prof. Acton and should include the Registration Fee of 10s.

MIND ASSOCIATION.

Those who wish to join the Association should communicate with the Hon. Secretary, Mrs. KNEALE, Lady Margaret Hall, Oxford; or with the Hon. Treasurer, Mr. H. STURT, 55 Park Town, Oxford, to whom the yearly subscription of sixteen shillings should be paid. Cheques should be made payable to the Mind Association, Westminster Bank, Oxford. Members may pay a Life Composition of £16 instead of the annual subscription.

In return for their subscriptions members receive MIND gratis and post free, and (if of 3 years' standing) are entitled to buy back-numbers of both the Old and the New Series at half-price.

Members resident in U.S.A. may pay the subscription ($4) to the Hon. Assistant-Treasurer, Prof. B. Blanshard, Swarthmore College, Swarthmore, Pennsylvania.

Members are reminded that subscriptions should be sent to the present Treasurer, Mr. Sturt, and *not* to Mr. McKie.